```
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGATGAGCGAAGTC 18
           | ||||||||||||||
Db      81 CACGATGAGCGAAGTC 96


RESULT  15
AV325088
LOCUS       AV325088      289 bp    mRNA            EST       11-NOV-1999
DEFINITION  AV325088 RIKEN full-length enriched, adult male medulla oblongata
            Mus musculus cDNA clone 6330400H24 3', mRNA sequence.
ACCESSION   AV325088
VERSION     AV325088.1  GI:6365140
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 289)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. a:  Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189274.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .289
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330400H24"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT       52 a    106 c     42 g     89 t
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 72;  Length 289;
  Best Local Similarity   93.8%;  Pred. No. 3.9e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 CCCCGATGAGCGAAGT 17
           |||| ||||||||||
Db     226 CCCCCATGAGCGAAGT 241


Search completed: June 22, 2000, 12:08:26
Job time: 10420 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:33 ; Search time 8627.09 Seconds
(without alignments)
-1.917 Million cell updates/sec

Title: US-09-362-485-16
Perfect score: 17
Sequence: 1 GGGGCCGTCCTGGTGCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	100.0	17	5	A87619	A87619 Sequence 16
2	17	100.0	720	5	A89759	A89759 Sequence 16
3	17	100.0	1194	2	MTU92472	U92472 Mycobacteri
4	17	100.0	1208	5	A89752	A89752 Sequence 9
5	17	100.0	1209	5	A87612	A87612 Sequence 9
6	17	100.0	1228	5	A87608	A87608 Sequence 5
7	17	100.0	1228	5	A89748	A89748 Sequence 5
8	17	100.0	1229	5	A87610	A87610 Sequence 7
9	17	100.0	1229	5	A89750	A89750 Sequence 7
10	17	100.0	1235	5	A87606	A87606 Sequence 3
11	17	100.0	1235	5	A87609	A87609 Sequence 6
12	17	100.0	1235	5	A87611	A87611 Sequence 8
13	17	100.0	1235	5	A89746	A89746 Sequence 3
14	17	100.0	1235	5	A89749	A89749 Sequence 6
15	17	100.0	1235	5	A89751	A89751 Sequence 8
16	17	100.0	1236	5	A87613	A87613 Sequence 10
17	17	100.0	1236	5	A89753	A89753 Sequence 10
18	17	100.0	1237	5	A87607	A87607 Sequence 4
19	17	100.0	1237	5	A89747	A89747 Sequence 4
20	17	100.0	1245	5	A87605	A87605 Sequence 2
21	17	100.0	1245	5	A89745	A89745 Sequence 2
22	17	100.0	1260	5	A87604	A87604 Sequence 1
23	17	100.0	1260	5	A89744	A89744 Sequence 1
24	17	100.0	2584	1	MTALADH	X63069 M.tuberculo
25	17	100.0	56414	1	MTV002	AL008967 Mycobacte
26	15.4	90.6	865	12	AF054894	AF054894 Rattus no
27	15.4	90.6	1447	12	RATMYOGEN	M24393 Rat myogeni
28	15.4	90.6	2589	12	AB019258	AB019258 Rattus no
29	15.4	90.6	2786	40	AF053719	AF053719 Homo sapi
30	15.4	90.6	2831	10	HSZ98265	Z98265 Homo sapien
31	15.4	90.6	40016	40	AC003038	AC003038 Human DNA
32	15.4	90.6	41055	1	SC6G4	AL031317 Streptomy
33	15.4	90.6	45399	33	HS305C8	AL031705 Homo sapi
c 34	15.4	90.6	110525	9	AB020878	AB020878 Homo sapi
35	15.4	90.6	117621	40	AC003950	AC003950 Homo sapi
c 36	15.4	90.6	121106	33	HSJ1004I9	AL121912 Homo sapi
c 37	15.4	90.6	124029	42	AC013560	AC013560 Homo sapi
38	15.4	90.6	128562	55	AC020962	AC020962 Mus muscu
39	15.4	90.6	141256	11	CNS01DUL	AL133304 Human chr
40	15.4	90.6	159107	32	AP001074	AP001074 Homo sapi
41	15.4	90.6	162591	33	AL133464	AL133464 Homo sapi
c 42	15.4	90.6	163253	43	AC011978	AC011978 Homo sapi
c 43	15.4	90.6	165418	52	AC020990	AC020990 Homo sapi
c 44	15.4	90.6	175203	45	AC022045	AC022045 Homo sapi
45	15.4	90.6	189805	40	AC004169	AC004169 Homo sapi

ALIGNMENTS

```
RESULT   1
A87619
LOCUS       A87619         17 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9836089.
ACCESSION   A87619
VERSION     A87619.1  GI:6736259
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        0 a      6 c      8 g      3 t
ORIGIN

  Query Match           100.0%;  Score 17;  DB 5;  Length 17;
  Best Local Similarity  100.0%;  Pred. No. 3.2e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db      1 GGGGCCGTCCTGGTGCC 17


RESULT   2
A89759
LOCUS       A89759         720 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9832862.
ACCESSION   A89759
VERSION     A89759.1  GI:6738291
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      133 a    238 c    226 g    123 t
ORIGIN

  Query Match           100.0%;  Score 17;  DB 5;  Length 720;
  Best Local Similarity  100.0%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    702 GGGGCCGTCCTGGTGCC 718


RESULT   3
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN

  Query Match           100.0%;  Score 17;  DB 2;  Length 1194;
  Best Local Similarity  100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT   4
A89752
LOCUS       A89752         1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT   5
A87612
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT   6
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT   7
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832852.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT   8
A87610
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from P tent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    768 GGGGCCGTCCTGGTGCC 784


RESULT   9
A89750
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
```

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    768 GGGGCCGTCCTGGTGCC 784


RESULT  10
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    769 GGGGCCGTCCTGGTGCC 785


RESULT  11
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    768 GGGGCCGTCCTGGTGCC 784


RESULT  12
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 17;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db    768 GGGGCCGTCCTGGTGCC 784


RESULT  13
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

Query Match 100.0%; Score 17; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GGGGCCGTCCTGGTGCC 17
           |||||||||||||||||
Db     769 GGGGCCGTCCTGGTGCC 785
```

RESULT 14
A89749
LOCUS A89749 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent WO9832862.
ACCESSION A89749
VERSION A89749.1 GI:6738283
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 235 a 395 c 384 g 220 t 1 others
ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GGGGCCGTCCTGGTGCC 17
           |||||||||||||||||
Db     768 GGGGCCGTCCTGGTGCC 784
```

RESULT 15
A89751
LOCUS A89751 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 8 from Patent WO9832862.
ACCESSION A89751
VERSION A89751.1 GI:6738285
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1235)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1235
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 394 c 385 g 220 t
ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GGGGCCGTCCTGGTGCC 17
           |||||||||||||||||
Db     768 GGGGCCGTCCTGGTGCC 784
```

Search completed: June 22, 2000, 14:58:35
Job time: 17930 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:48 ; Search time 458.59 Seconds
(without alignments)
9.275 Million cell updates/sec

Title: US-09-362-485-16
Perfect score: 17
Sequence: 1 GGGGCCGTCCTGGTGCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		17	100.0	17	1	V49517	Mycobacterium sp.
2		17	100.0	17	1	V49615	AlaDH derived olig
3		17	100.0	1245	1	V49626	Mycobacterium tube
4		17	100.0	1260	1	V49510	Mycobacterium sp.
5		17	100.0	1260	1	V49625	Mycobacterium tube
6		15.4	90.6	329	1	V89242	EST clone C1126. N
7		15	88.2	684	1	V84485	Human secreted pro
8		15	88.2	24379	1	T93095	Streptomyces freno
9		15	88.2	24379	1	V25925	Streptomyces roseo
10		14.4	84.7	1575	1	T30361	Bovine p57 coding
c 11		14.4	84.7	1590	1	Q05246	Sequence encoding
c 12		14.4	84.7	1650	1	V37926	Platelet derived e
13		14.4	84.7	1657	1	T30360	Human p57 coding s
14		14.4	84.7	1842	1	T30370	GST/truncated huma
15		14.4	84.7	2067	1	T30369	GST/human p57 fusi
16		14.4	84.7	3632	1	T06978	T. thermophilus ga
17		14.4	84.7	3632	1	X21373	Thermus thermophil
18		14.4	84.7	3870	1	T12292	Phospholipase C-ga
19		14.4	84.7	3893	1	T12293	Phospholipase C-ga
c 20		14.4	84.7	4342	1	T16470	Murine hepatoma tr
c 21		14.4	84.7	4698	1	Q32370	P1A genomic DNA. N
c 22		14.4	84.7	4698	1	Q72475	Entire P1A genomic
c 23		14.4	84.7	4698	1	T01176	P815A antigen prec
c 24		14.4	84.7	4698	1	T14692	MAGE genomic DNA.
25		14	82.4	977	1	N20007	Hybrid human leuko
26		14	82.4	4344	1	X07582	Equus caballus Y c
c 27		14	82.4	6743	1	X32515	T-DNA containing t
c 28		13.8	81.2	45	1	Q20698	Immunostimulatory
c 29		13.8	81.2	45	1	Q20860	Immunostimulatory
c 30		13.8	81.2	45	1	Q20869	Immunostimulatory
c 31		13.8	81.2	60	1	T44506	Human MUC1 gene ta
c 32		13.8	81.2	60	1	T38664	Human MUC1 gene ta
c 33		13.8	81.2	60	1	T88166	Human MUC1 tandem
c 34		13.8	81.2	60	1	V48316	Nucleotide sequenc
c 35		13.8	81.2	180	1	T29189	Glycoprotein 39 5'
c 36		13.8	81.2	268	1	Q57761	Mycobacterium bovi
c 37		13.8	81.2	268	1	Q57762	Mycobacterium micr
c 38		13.8	81.2	268	1	Q57763	Mycobacterium afri
39		13.8	81.2	309	1	N90579	pDF9.3 cDNA insert
c 40		13.8	81.2	342	1	Q06507	Encodes protein IV
41		13.8	81.2	345	1	X24395	Mouse DNAX accesso
42		13.8	81.2	452	1	V83516	Killer cell activa
c 43		13.8	81.2	508	1	V45133	Homo sapiens cerbe
44		13.8	81.2	603	1	V84058	cDNA encoding a cy
c 45		13.8	81.2	804	1	X08989	Sequence encoding

ALIGNMENTS

RESULT 1
V49517
ID V49517 standard; DNA; 17 BP.
AC V49517;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F6.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 17 BP; 0 A; 6 C; 8 G; 3 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||||||||||||||||
Db 1 GGGGCCGTCCTGGTGCC 17

RESULT 2
V49615
ID V49615 standard; DNA; 17 BP.
AC V49615;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F6.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 17 BP; 0 A; 6 C; 8 G; 3 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||||||||||||||||
Db 1 GGGGCCGTCCTGGTGCC 17

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 17; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||||||||||||||||
Db 775 GGGGCCGTCCTGGTGCC 791

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 17; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||||||||||||||||
Db 790 GGGGCCGTCCTGGTGCC 806

RESULT 5
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 17; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||||||||||||||||
Db 790 GGGGCCGTCCTGGTGCC 806

RESULT 6
V89242
ID V89242 standard; cDNA; 329 BP.
AC V89242;
DT 15-FEB-1999 (first entry)
DE EST clone C1126.
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN WO9845436-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070077/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 155; 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
SQ Sequence 329 BP; 61 A; 78 C; 142 G; 48 T;

Query Match 90.6%; Score 15.4; DB 1; Length 329;
Best Local Similarity 94.1%; Pred. No. 56;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGCC 17
||||| ||||||||||
Db 131 GGGGCGGTCCTGGTGCC 147

RESULT 7
V84485
ID V84485 standard; DNA; 684 BP.
AC V84485;
DT 01-MAR-1999 (first entry)
DE Human secreted protein gene 75 clone HJPAV06.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9854963-A2.
PD 10-DEC-1998.
PF 04-JUN-1998; U11422.
PR 18-DEC-1997; US-070923.
PR 06-JUN-1997; US-048877.
PR 06-JUN-1997; US-048881.
PR 06-JUN-1997; US-048884.
PR 06-JUN-1997; US-048893.
PR 06-JUN-1997; US-048896.
PR 06-JUN-1997; US-048899.
PR 06-JUN-1997; US-048915.
PR 06-JUN-1997; US-048949.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048972.
PR 06-JUN-1997; US-049020.
PR 06-JUN-1997; US-049375.
PR 05-SEP-1997; US-057628.
PR 05-SEP-1997; US-057635.
PR 05-SEP-1997; US-057644.
PR 05-SEP-1997; US-057647.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057661.
PR 05-SEP-1997; US-057667.
PR 05-SEP-1997; US-057761.
PR 05-SEP-1997; US-057764.
PR 05-SEP-1997; US-057770.
PR 05-SEP-1997; US-057775.
PR 05-SEP-1997; US-057778.
PR 06-JUN-1997; US-048875.
PR 06-JUN-1997; US-048878.
PR 06-JUN-1997; US-048882.
PR 06-JUN-1997; US-048885.
PR 06-JUN-1997; US-048894.
PR 06-JUN-1997; US-048897.
PR 06-JUN-1997; US-048900.
PR 06-JUN-1997; US-048916.
PR 06-JUN-1997; US-048962.
PR 06-JUN-1997; US-048970.
PR 06-JUN-1997; US-048974.
PR 06-JUN-1997; US-049373.
PR 05-SEP-1997; US-057584.
PR 05-SEP-1997; US-057629.
PR 05-SEP-1997; US-057642.
PR 05-SEP-1997; US-057645.
PR 05-SEP-1997; US-057648.
PR 05-SEP-1997; US-057651.
PR 05-SEP-1997; US-057662.
PR 05-SEP-1997; US-057668.
PR 05-SEP-1997; US-057762.
PR 05-SEP-1997; US-057765.
PR 05-SEP-1997; US-057771.
PR 05-SEP-1997; US-057776.
PR 06-JUN-1997; US-048876.
PR 06-JUN-1997; US-048880.
PR 06-JUN-1997; US-048883.
PR 06-JUN-1997; US-048892.
PR 06-JUN-1997; US-048895.
PR 06-JUN-1997; US-048898.
PR 06-JUN-1997; US-048901.
PR 06-JUN-1997; US-048917.
PR 06-JUN-1997; US-048963.
PR 06-JUN-1997; US-048971.
PR 06-JUN-1997; US-049019.
PR 06-JUN-1997; US-049374.

PR 05-SEP-1997; US-057627.
PR 05-SEP-1997; US-057634.
PR 05-SEP-1997; US-057643.
PR 05-SEP-1997; US-057646.
PR 05-SEP-1997; US-057649.
PR 05-SEP-1997; US-057654.
PR 05-SEP-1997; US-057666.
PR 05-SEP-1997; US-057760.
PR 05-SEP-1997; US-057763.
PR 05-SEP-1997; US-057769.
PR 05-SEP-1997; US-057774.
PR 05-SEP-1997; US-057777.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR WPI; 99-059865/05.
DR P-PSDB; W88608, W88840.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 4; Page 339; 772pp; English.
CC The invention relates to nucleic acid sequences (V84411 to V84633)
CC encoding human secreted proteins (W88534 to W88756). The secreted protein
CC gene sequences are deposited with the ATCC under deposit numbers ATCC
CC 97979, 97974, 97975, 97976, 97977, 209007, 209018, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
SQ Sequence 684 BP; 131 A; 238 C; 170 G; 142 T;

Query Match 88.2%; Score 15; DB 1; Length 684;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTG 15
||||||||||||||| |
Db 210 GGGGCCGTCCTGGTG 224

RESULT 8
T93095
ID T93095 standard; cDNA; 24379 BP.
AC T93095;
DT 11-MAY-1998 (first entry)
DE Streptomyces frenolicin gene cluster.
KW Frenolicin; antibiotic; feed additive; anticoccidial;
KW coccidiostatic; efflux pump; butyrate starter synthase;
KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;
KW ds.
OS Streptomyces sp.

FH	Key	Location/Qualifiers
FT	CDS	636. .2948
FT		/*tag= a
FT		/product= 30 kDa non-membrane protein
FT		/note= "gene A (specifically claimed)"
FT	CDS	2945. .3916
FT		/*tag= b
FT		/product= membrane protein
FT		/note= "gene B (specifically claimed)"
FT	CDS	4020. .4844
FT		/*tag= c
FT		/product= protein with 6 membrane-spanning domains
FT		/note= "gene C (specifically claimed)"
FT	CDS	4841. .6415
FT		/*tag= d
FT		/product= ATP-binding component of ABC transporter
FT		/note= "gene D (specifically claimed)"
FT	CDS	6533. .7187
FT		/*tag= e
FT		/product= unknown non-membrane protein
FT		/note= "gene E (specifically claimed)"
FT	CDS	7344. .8897
FT		/*tag= f
FT		/product= putative efflux pump
FT		/note= "gene F (specifically claimed)"
FT	CDS	9164. .10012
FT		/*tag= g
FT		/product= transcription activator
FT		/note= "gene G (specifically claimed)"
FT	CDS	complement (10105. .10621)
FT		/*tag= h
FT		/product= translationally coupled to gene I
FT		/note= "gene H (specifically claimed)"
FT	CDS	complement (10618. .11628)
FT		/*tag= i
FT		/product= homologue of fabH
FT		/note= "gene I (specifically claimed)"
FT	CDS	11809. .12066
FT		/*tag= j
FT		/product= acyl carrier protein
FT		/note= "gene J (specifically claimed)"
FT	CDS	complement (12154. .13209)
FT		/*tag= k
FT		/product= acyltransferase
FT		/note= "gene K (specifically claimed)"
FT	CDS	13409. .14686
FT		/*tag= l
FT		/product= PKS ketoacylsynthase subunit
FT		/note= "gene L (specifically claimed)"
FT	CDS	14767. .16047
FT		/*tag= m
FT		/product= PKS chain length factor
FT		/note= "gene M (specifically claimed)"
FT	CDS	16120. .16371
FT		/*tag= n
FT		/product= acyl carrier protein used by the PKS
FT		/note= "gene N (specifically claimed)"
FT	CDS	complement (16453. .16935)
FT		/*tag= o
FT		/product= putative hemiketal dehydrase
FT		/note= "gene O (specifically claimed)"
FT	CDS	17088. .17903
FT		/*tag= p
FT		/product= ketoreductase related to actIII
FT		/note= "gene P (specifically claimed)"
FT	CDS	17903. .18898
FT		/*tag= q
FT		/product= cyclase/dehydrase related to act VII
FT		/note= "gene Q (specifically claimed)"
FT	CDS	18895. .19839
FT		/*tag= r
FT		/product= cyclase/dehydrase related to actIV
FT		/note= "gene R (specifically claimed)"

```
FT   CDS             complement (19990. .20907)
FT                   /*tag= s
FT                   /product= oxidoreductase
FT                   /note= "gene S (specifically claimed)"
FT   CDS             complement (20904. .22094)
FT                   /*tag= t
FT                   /product= quinone-forming hydroxylase
FT                   /note= "gene T (specifically claimed)"
FT   CDS             complement (22505. .22179)
FT                   /*tag= u
FT                   /product= not specified
FT                   /note= "gene U"
PN   EP-806480-A2.
PD   12-NOV-1997.
PF   02-MAY-1997; 107329.
PR   07-MAY-1996; US-016753.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Reeves CD, Soliday CL;
DR   WPI; 97-538619/50.
DR   P-PSDB; W34199-W34219.
PT   Streptomyces frenolicin gene cluster - useful for producing
PT   recombinant frenolicin antibiotics
PS   Claim 1; Page 40-60; 66pp; English.
CC   This DNA sequence comprises the Streptomyces frenolicin gene
CC   cluster containing specifically claimed coding sequences (genes
CC   A-U) that respectively encode 21 proteins (see W34199-219) involved
CC   in frenolicin synthesis.  The genes can be divided into 5
CC   subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC   genes H, I, J and K encode butyrate starter synthases; (3) genes
CC   L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC   R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC   and (5) genes S and T encode a keto/enoyl reductase and a
CC   hydrolase.  Also claimed are vectors, host cells (especially a
CC   Streptomyces sp., particularly Streptomyces roseofulvus), and the
CC   encoded proteins.  Cells transformed using the above sequence can
CC   be cultured to produce frenolicins or frenolicin precursors.  The
CC   precursors can be converted to frenolicins by chemical or other
CC   methods.  The frenolicins can be oxidised to frenolicin B, an
CC   antibiotic used as an anticoccidial agent.  The frenolicins can be
CC   used as animal feed additives.
SQ   Sequence   24379 BP;    3077 A;    9792 C;    8499 G;    3011 T;


  Query Match             88.2%;  Score 15;  DB 1;  Length 24379;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTG 15
           |||||||||||||||
Db    3704 GGGGCCGTCCTGGTG 3718


RESULT   9
V25925
ID   V25925 standard; cDNA; 24379 BP.
AC   V25925;
DT   15-JUL-1998  (first entry)
DE   Streptomyces roseofulvus frenolicin gene cluster.
KW   Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
KW   antibiotic; ss.
OS   Streptomyces roseofulvus.
FH   Key             Location/Qualifiers
FT   CDS             636. .2948
FT                   /*tag= a
FT                   /transl_except= (pos:636. .638,aa:Met)
FT                   /note= "encodes protein given in W55800"
FT   CDS             2945. .3916
FT                   /*tag= b
FT                   /transl_except= (pos:2945. .2947,aa:Met)
FT                   /note= "encodes protein given in W55801"
FT   CDS             4020. .4844
FT                   /*tag= c
FT                   /transl_except= (pos:4020. .4022,aa:Met)
FT                   /note= "encodes protein given in W55802"
FT   CDS             4841. .6415
FT                   /*tag= d
FT                   /note= "encodes protein given in W55803"
FT   CDS             6533. .7183
FT                   /*tag= e
FT                   /transl_except= (pos:6533. .6535,aa:Met)
FT                   /note= "encodes protein given in W55804"
FT   CDS             7344. .8897
FT                   /*tag= f
FT                   /note= "encodes protein given in W55805"
FT   CDS             9164. .10012
FT                   /*tag= g
FT                   /transl_except= (pos:9164. .9166,aa:Met)
FT                   /note= "encodes protein given in W55806"
FT   CDS             complement (10105. .10621)
FT                   /*tag= h
FT                   /transl_except= (pos:10621. .10619,aa:Met)
FT                   /note= "encodes protein given in W55807"
FT   CDS             complement (10618. .11628)
FT                   /*tag= i
FT                   /transl_except= (pos:11628. .11626,aa:Met)
FT                   /note= "encodes protein given in W55808"
FT   CDS             11809. .12066
FT                   /*tag= j
FT                   /note= "encodes protein given in W55809"
FT   CDS             complement (12154. .13209)
FT                   /*tag= k
FT                   /transl_except= (pos:13209. .13207,aa:Met)
FT                   /note= "encodes protein given in W55810"
FT   CDS             13409. .14686
FT                   /*tag= l
FT                   /transl_except= (pos:13409. .13411,aa:Met)
FT                   /note= "encodes protein given in W55811"
FT   CDS             14767. .16047
FT                   /*tag= m
FT                   /note= "encodes protein given in W55812"
FT   CDS             16120. .16371
FT                   /*tag= n
FT                   /note= "encodes protein given in W55813"
FT   CDS             complement (16453. .16935)
FT                   /*tag= o
FT                   /transl_except= (pos:16935. .16933,aa:Met)
FT                   /note= "encodes protein given in W55814"
FT   CDS             17088. .18903
FT                   /*tag= p
FT                   /note= "encodes protein given in W55815"
FT   CDS             17903. .18898
FT                   /*tag= q
FT                   /note= "encodes protein given in W55816"
FT   CDS             18895. .19839
FT                   /*tag= r
FT                   /note= "encodes protein given in W55817"
FT   CDS             complement (19990. .20907)
FT                   /*tag= s
FT                   /transl_except= (pos:20907. .20905,aa:Met)
FT                   /note= "encodes protein given in W55818"
FT   CDS             complement (20904. .22094)
FT                   /*tag= t
FT                   /transl_except= (pos:22040. .22038,aa:Arg)
FT                   /transl_except= (pos:22037. .22035,aa:Arg)
FT                   /transl_except= (pos:22034. .22032,aa:Thr)
FT                   /transl_except= (pos:22031. .22029,aa:Lys)
FT                   /note= "encodes protein given in W55819"
PN   J10094395-A.
PD   14-APR-1998.
PF   07-MAY-1997; 116652.
PR   04-APR-1997; US-042935.
PR   07-MAY-1996; US-016753.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
DR   WPI; 98-279231/25.
DR   P-PSDB; W55800-W55819.
PT   Frenolicin gene cluster - useful as an antibiotic
```

```
PS   Claim 2; Page 36-44; 50pp; Japanese.
CC   The present sequence represents a frenolicin gene cluster from
CC   Streptomyces roseofulvus. The present invention describes: (1) a
CC   vector containing the frenolicin gene cluster DNA sequence operably
CC   connected to an expression control sequence; (2) a host cell
CC   transformed by the above vector; (3) a protein coded by the above DNA
CC   sequence; (4) a method for the preparation of frenolicin or a
CC   biosynthetic intermediate for it in which the above cell is cultured
CC   and frenolicin or its biosynthetic intermediate is isolated from the
CC   culture or the cell; (5) a method for the preparation of frenolicin B
CC   by oxidising frenolicin, and (6) a method for the preparation of a
CC   feed composition by mixing frenolicin with other components.
CC   Frenolicin B is useful as an antibiotic.
SQ   Sequence   24379 BP;    3077 A;    9792 C;    8499 G;    3011 T;

  Query Match             88.2%;  Score 15;  DB 1;  Length 24379;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTG 15
          |||||||||||||||
Db   3704 GGGGCCGTCCTGGTG 3718


RESULT  10
T30361
ID   T30361 standard; cDNA; 1575 BP.
AC   T30361;
DT   26-NOV-1996  (first entry)
DE   Bovine p57 coding sequence.
KW   WD40 repeated structure region; bovine p57 protein; dimer;
KW   anti-cancer; anti-HIV agent; actin; cell movement; ds.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   cds             93. .1478
FT                   /*tag= a
FT                   /product= p57
PN   J08119996-A.
PD   14-MAY-1996.
PF   21-OCT-1994; 282743.
PR   21-OCT-1994; JP-282743.
PA   (NISB ) JAPAN TOBACCO INC.
DR   WPI; 96-283507/29.
DR   P-PSDB; R98342.
PT   A new protein, p57, comprising WD40 repeat region - used for
PT   development of anti-cancer and anti-HIV agents
PS   Example 2; Page 27-29; 51pp; Japanese.
CC   This sequence encodes the bovine p57 protein.  p57 contains a leucine
CC   rich C-terminal peptide which comprises a leucine residue after each
CC   7 amino acids and a  WD40 repeated structure region containing five
CC   WD40 regions.  p57 forms a dimer.  It can be used in the development
CC   of an anti-cancer agent and an anti-HIV agent.  It can also combine
CC   with actin in the control of cell movement.
SQ   Sequence   1575 BP;    304 A;    482 C;    492 G;    297 T;

  Query Match             84.7%;  Score 14.4;  DB 1;  Length 1575;
  Best Local Similarity   93.8%;  Pred. No. 1.5e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCC.GGTGC 16
          ||||||.|||||||||
Db    261 GGGGCCTTCCTGGTGC 276


RESULT  11
Q05246/c
ID   Q05246 standard; DNA; 1590 BP.
AC   Q05246;
DT   03-DEC-1990  (first entry)
DE   Sequence encoding placental platelet-derived endothelial cell
DE   growth factor (PD-ECGF).
KW   Platelet-derived endothelial cell growth factor; PD-ECGF;
KW   angiogenisis; atherosclerosis; thrombocytopenia; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             124. .1569
FT                   /*tag= a
PN   EP-377855-A.
PD   18-JUL-1990.
PF   15-DEC-1989; 123229.
PR   20-DEC-1988; US-288056.
PA   (RESE) Research Corp Techn.
PI   Heldin CH, Kiyazono K, Wernstedt C, Hellman U, Takaku F;
PI   Ishikawa F.;
DR   WPI; 90-218332/29.
DR   P-PSDB; R05897.
PT   Recombinant platelet-derived endothelial cell growth factor -
PT   used for inducing angiogenisis or treating wounds, atherosclerosis
PT   or thrombocytopenia.
PS   Claim 17; Fig 11B; 49pp; English.
CC   Polypeptide gene product can be produced in a microorganism or
CC   cell culture expression system, using a sequence derived from
CC   placental DNA. PD-ECGF is useful in treating wounds, atherosclerosis,
CC   and thrombocytopenia, or to induce angiogenesis. Abs raised to the
CC   product may be used to locate cells or tissues expressing it.
SQ   Sequence   1590 BP;    249 A;    511 C;    593 G;    237 T;

  Query Match             84.7%;  Score 14.4;  DB 1;  Length 1590;
  Best Local Similarity   93.8%;  Pred. No. 1.5e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          |||||||||| |||||
Db   1458 GGGGCCG-CCCGGTGC 1443


RESULT  12
V37926/c
ID   V37926 standard; cDNA; 1650 BP.
AC   V37926;
DT   10-SEP-1998  (first entry)
DE   Platelet derived endothelial cell growth factor encoding cDNA.
KW   Recombinant; platelet derived; endothelial cell growth factor;
KW   PD-ECGF; human; mitogen; angiogenic; chemotaxis; atherosclerosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             124. .1632
FT                   /*tag= a
FT                   /product= "PD-ECGF"
PN   US5756686-A.
PD   26-MAY-1998.
PF   15-NOV-1991; 192757.
PR   15-NOV-1991; US-792891.
PR   20-DEC-1988; US-288056.
PR   16-OCT-1989; US-422379.
PR   15-NOV-1991; US-192757.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Heldin C, Hellman U, Ishikawa F, Miyazono K, Takaku F,
PI   Wernstedt C;
DR   WPI; 98-321633/28.
DR   P-PSDB; W62026.
PT   Recombinant platelet-derived endothelial cell growth factor and
PT   peptide fragments - useful to treat atherosclerosis, to promote
PT   wound healing and to induce angiogenesis and specific chemotaxis of
PT   endothelial cells
PS   Example; Column 9-12; 41pp; English.
CC   The present sequence encodes an isolated recombinantly produced
CC   platelet derived endothelial cell growth factor (PD-ECGF). The ECGF is
CC   capable of inducing an endothelial cell chemotactic response. The ECGF
CC   has a molecular weight of 45 kD, does not bind to heparin, does not
CC   stimulate fibroblast proliferation, and is angiogenic. Recombinant
CC   PD-ECGF can be used to treat atherosclerosis, to promote wound healing
CC   and to induce angiogenesis and specific chemotaxis of endothelial cells.
```

SQ Sequence 1650 BP; 258 A; 547 C; 601 G; 244 T;

Query Match 84.7%; Score 14.4; DB 1; Length 1650;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGC 16
||||||||||| |||||
Db 1518 GGGGCCGTCCCGGTGC 1503

RESULT 13
T30360
ID T30360 standard; cDNA; 1657 BP.
AC T30360;
DT 26-NOV-1996 (first entry)
DE Human p57 coding sequence.
KW WD40 repeated structure region; human p57 protein; dimer;
KW anti-cancer; anti-HIV agent; actin; cell movement; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 100. .1485
FT /*tag= a
FT /product= p57
PN J08119996-A.
PD 14-MAY-1996.
PF 21-OCT-1994; 282743.
PR 21-OCT-1994; JP-282743.
PA (NISB) JAPAN TOBACCO INC.
DR WPI; 96-283507/29.
DR P-PSDB; R98341.
PT A new protein, p57, comprising WD40 repeat region - used for
PT development of anti-cancer and anti-HIV agents
PS Claim 9; Page 23-25; 51pp; Japanese.
CC This sequence encodes the human p57 protein. p57 contains a leucine
CC rich C-terminal peptide which comprises a leucine residue after
CC each 7 amino acids (see also R983340) and a WD40 repeated structure
CC region (see also R98334-38) containing five WD40 regions. p57 forms a
CC dimer. It can be used in the development of an anti-cancer agent and
CC an anti-HIV agent. It can also combine with actin in the control of
CC cell movement.
SQ Sequence 1657 BP; 359 A; 497 C; 511 G; 290 T;

Query Match 84.7%; Score 14.4; DB 1; Length 1657;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGC 16
|||||| |||||||||
Db 268 GGGGCCTTCCTGGTGC 283

RESULT 14
T30370
ID T30370 standard; cDNA; 1842 BP.
AC T30370;
DT 26-NOV-1996 (first entry)
DE GST/truncated human p57 fusion protein coding sequence.
KW WD40 repeated structure region; bovine p57 protein; dimer;
KW anti-cancer; anti-HIV agent; actin; cell movement; ds.
OS Synthetic.
PN J08119996-A.
PD 14-MAY-1996.
PF 21-OCT-1994; 282743.
PR 21-OCT-1994; JP-282743.
PA (NISB) JAPAN TOBACCO INC.
DR WPI; 96-283507/29.
DR P-PSDB; R98344.
PT A new protein, p57, comprising WD40 repeat region - used for
PT development of anti-cancer and anti-HIV agents
PS Disclosure; Page 33-35; 51pp; Japanese.
CC This sequence encodes a glutathione S transferase/truncated human p57
CC fusion protein. This sequence is derived from a derivative of the
CC plasmid pGEX/hp57. This plasmid is used in the expression of the
CC human p57 protein. p57 contains a leucine rich C-terminal peptide
CC which comprises a leucine residue after each 7 amino acids and a
CC WD40 repeated structure region containing five WD40 regions. p57
CC forms a dimer. It can be used in the development of an anti-cancer
CC agent and an anti-HIV agent. It can also combine with actin in the
CC control of cell movement.
SQ Sequence 1842 BP; 422 A; 462 C; 532 G; 426 T;

Query Match 84.7%; Score 14.4; DB 1; Length 1842;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGC 16
|||||| |||||||||
Db 853 GGGGCCTTCCTGGTGC 868

RESULT 15
T30369
ID T30369 standard; cDNA; 2067 BP.
AC T30369;
DT 26-NOV-1996 (first entry)
DE GST/human p57 fusion protein coding sequence.
KW WD40 repeated structure region; bovine p57 protein; dimer;
KW anti-cancer; anti-HIV agent; actin; cell movement; ds.
OS Synthetic.
PN J08119996-A.
PD 14-MAY-1996.
PF 21-OCT-1994; 282743.
PR 21-OCT-1994; JP-282743.
PA (NISB) JAPAN TOBACCO INC.
DR WPI; 96-283507/29.
DR P-PSDB; R98343.
PT A new protein, p57, comprising WD40 repeat region - used for
PT development of anti-cancer and anti-HIV agents
PS Example 5; Page 30-32; 51pp; Japanese.
CC This sequence encodes a glutathione S transferase/human p57 fusion
CC protein. This sequence is derived from the plasmid pGEX/hp57. This
CC plasmid is used in the expression of the human p57 protein. p57
CC contains a leucine rich C-terminal peptide which comprises a leucine
CC residue after each 7 amino acids and a WD40 repeated structure region
CC containing five WD40 regions. p57 forms a dimer. It can be used
CC in the development of an anti-cancer agent and an anti-HIV agent.
CC It can also combine with actin in the control of cell movement.
SQ Sequence 2067 BP; 470 A; 532 C; 612 G; 453 T;

Query Match 84.7%; Score 14.4; DB 1; Length 2067;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGGCCGTCCTGGTGC 16
|||||| |||||||||
Db 853 GGGGCCTTCCTGGTGC 868

Search completed: June 22, 2000, 15:12:51
Job time: 18097 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:57 ; Search time 341.15 Seconds
(without alignments)
6.477 Million cell updates/sec

Title: US-09-362-485-16
Perfect score: 17
Sequence: 1 GGGGCCGTCCTGGTGCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	14.4	84.7	3870	1	US-08-138-641-1	Sequence 1, Appli
	2	14.4	84.7	3870	1	US-08-138-133-1	Sequence 1, Appli
	3	14.4	84.7	3893	1	US-08-138-641-3	Sequence 3, Appli
	4	14.4	84.7	3893	1	US-08-138-133-3	Sequence 3, Appli
c	5	14.4	84.7	4342	1	US-08-436-044-1	Sequence 1, Appli
c	6	14.4	84.7	4342	3	US-08-436-054-1	Sequence 1, Appli
c	7	14.4	84.7	4342	6	PCT-US95-08812-1	Sequence 1, Appli
c	8	14.4	84.7	4698	1	US-07-807-043B-5	Sequence 5, Appli
c	9	14.4	84.7	4698	1	US-08-299-849B-5	Sequence 5, Appli
c	10	14.4	84.7	4698	4	US-08-142-368A-5	Sequence 5, Appli
c	11	14.4	84.7	4698	5	US-08-967-727-5	Sequence 5, Appli
c	12	14	82.4	6743	5	US-08-932-280-1	Sequence 1, Appli
c	13	13.8	81.2	268	1	US-08-105-168B-1	Sequence 1, Appli
c	14	13.8	81.2	268	1	US-08-105-168B-2	Sequence 2, Appli
c	15	13.8	81.2	268	1	US-08-105-168B-3	Sequence 3, Appli
c	16	13.8	81.2	268	3	US-08-698-948-1	Sequence 1, Appli
c	17	13.8	81.2	268	3	US-08-698-948-2	Sequence 2, Appli
c	18	13.8	81.2	268	3	US-08-698-948-3	Sequence 3, Appli
c	19	13.8	81.2	314	1	US-08-105-168B-23	Sequence 23, Appl
c	20	13.8	81.2	314	3	US-08-698-948-23	Sequence 23, Appl
c	21	13.8	81.2	342	3	US-08-473-020A-31	Sequence 31, Appl
c	22	13.8	81.2	343	3	US-08-473-020A-4	Sequence 4, Appli
c	23	13.8	81.2	350	1	US-08-105-168B-22	Sequence 22, Appl
c	24	13.8	81.2	350	3	US-08-698-948-22	Sequence 22, Appl
c	25	13.8	81.2	1800	3	US-08-733-825-1	Sequence 1, Appli
	26	13.8	81.2	2582	4	US-08-816-105A-2	Sequence 2, Appli
	27	13.8	81.2	3404	1	US-07-762-132A-1	Sequence 1, Appli
	28	13.8	81.2	7498	3	US-08-816-693A-1	Sequence 1, Appli
	29	13.8	81.2	11558	6	PCT-US93-06251-23	Sequence 23, Appl
c	30	13.8	81.2	43280	3	US-08-804-227C-1	Sequence 1, Appli
	31	13.4	78.8	537	1	US-08-176-620A-9	Sequence 9, Appli
	32	13.4	78.8	537	3	US-08-461-985-9	Sequence 9, Appli
	33	13.4	78.8	847	3	US-08-773-423-4	Sequence 4, Appli
	34	13.4	78.8	1560	4	US-08-794-795-5	Sequence 5, Appli
	35	13.4	78.8	1703	4	US-08-794-795-1	Sequence 1, Appli
	36	13.4	78.8	1776	2	US-08-484-840-1	Sequence 1, Appli
	37	13.4	78.8	1776	2	US-08-483-094-1	Sequence 1, Appli
	38	13.4	78.8	2085	2	US-08-553-999B-1	Sequence 1, Appli
	39	13.4	78.8	2911	5	US-08-981-825-5	Sequence 5, Appli
c	40	13.4	78.8	3836	3	US-08-216-260-1	Sequence 1, Appli
	41	13.4	78.8	246240	3	US-08-724-394A-20	Sequence 20, Appl
	42	13.4	78.8	246240	3	US-08-724-394A-21	Sequence 21, Appl
	43	13.4	78.8	246240	3	US-08-724-394A-22	Sequence 22, Appl
c	44	13	76.5	1513	1	US-07-918-314-1	Sequence 1, Appli
	45	13	76.5	2259	7	5185254-3	Patent No. 5185254

ALIGNMENTS

```
RESULT   1
US-08-138-641-1
; Sequence 1, Application US/08138641
; Patent No. 5474921
;  GENERAL INFORMATION:
;    APPLICANT:  Koblan, Kenneth S.
;    APPLICANT:  Pompliano, David L.
;    TITLE OF INVENTION:  ASSAY TO DETERMINE INHIBITORS OF
;    TITLE OF INVENTION:  PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C-GAMMA
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Muthard
;      STREET:  P.O. Box 2000, 126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,641
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Muthard, David A.
;      REGISTRATION NUMBER:  35,297
;      REFERENCE/DOCKET NUMBER:  18937
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3903
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3870 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-138-641-1


  Query Match             84.7%;  Score 14.4;  DB 1;  Length 3870;
  Best Local Similarity   93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           |||||| |||||||||
Db    2065 GGGGCCTTCCTGGTGC 2080
```

```
RESULT   2
US-08-138-133-1
; Sequence 1, Application US/08138133
; Patent No. 5519163
;  GENERAL INFORMATION:
;    APPLICANT:  GIBBS, JACKSON B.
;    APPLICANT:  KOBLAN, KENNETH S.
;    APPLICANT:  MACLEOD, ANGUS M.
;    APPLICANT:  MERCHANT, KEVIN J.
;    TITLE OF INVENTION:  INHIBITORS OF PHOSPHOINOSITIDE-SPECIFIC
;    TITLE OF INVENTION:  PHOSPHOLIPASE C
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID A. MUTHARD
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NEW JERSEY
;      COUNTRY:  U.S.A.
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,133
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MUTHARD, DAVID A.
;      REGISTRATION NUMBER:  35,297
;      REFERENCE/DOCKET NUMBER:  18938
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3903
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3870 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-138-133-1


  Query Match           84.7%;  Score 14.4;  DB 1;  Length 3870;
  Best Local Similarity  93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          |||||| |||||||||
Db   2065 GGGGCCTTCCTGGTGC 2080


RESULT   3
US-08-138-641-3
; Sequence 3, Application US/08138641
; Patent No. 5474921
;  GENERAL INFORMATION:
;    APPLICANT:  Koblan, Kenneth S.
;    APPLICANT:  Pompliano, David L.
;    TITLE OF INVENTION:  ASSAY TO DETERMINE INHIBITORS OF
;    TITLE OF INVENTION:  PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C-GAMMA
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Muthard
;      STREET:  P.O. Box 2000, 126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,641
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Muthard, David A.
;      REGISTRATION NUMBER:  35,297
;      REFERENCE/DOCKET NUMBER:  18937
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3903
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3893 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-138-641-3


  Query Match           84.7%;  Score 14.4;  DB 1;  Length 3893;
  Best Local Similarity  93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          |||||| |||||||||
Db   2065 GGGGCCTTCCTGGTGC 2080


RESULT   4
US-08-138-133-3
; Sequence 3, Application US/08138133
; Patent No. 5519163
;  GENERAL INFORMATION:
;    APPLICANT:  GIBBS, JACKSON B.
;    APPLICANT:  KOBLAN, KENNETH S.
;    APPLICANT:  MACLEOD, ANGUS M.
;    APPLICANT:  MERCHANT, KEVIN J.
;    TITLE OF INVENTION:  INHIBITORS OF PHOSPHOINOSITIDE-SPECIFIC
;    TITLE OF INVENTION:  PHOSPHOLIPASE C
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID A. MUTHARD
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NEW JERSEY
;      COUNTRY:  U.S.A.
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,133
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MUTHARD, DAVID A.
;      REGISTRATION NUMBER:  35,297
;      REFERENCE/DOCKET NUMBER:  18938
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 594-3903
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3893 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-138-133-3


  Query Match             84.7%;  Score 14.4;  DB 1;  Length 3893;
  Best Local Similarity   93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          ||||||| |||||||||
Db   2065 GGGGCCTTCCTGGTGC 2080


RESULT   5
US-08-436-044-1/c
; Sequence 1, Application US/08436044
; Patent No. 5624899
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Matthews, William
;    TITLE OF INVENTION:  HTK LIGAND
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/436,044
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/277722
;      FILING DATE:  20-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  902D3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4342 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-436-044-1


  Query Match             84.7%;  Score 14.4;  DB 1;  Length 4342;
  Best Local Similarity   93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          ||||||| |||||||||
Db   1503 GGGGCCGGCCTGGTGC 1488


RESULT   6
US-08-436-054-1/c
; Sequence 1, Application US/08436054
; Patent No. 5864020
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Matthews, William
;    TITLE OF INVENTION:  HTK LIGAND
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/436,054
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/277722
;      FILING DATE:  20-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  902D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4342 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-436-054-1


  Query Match             84.7%;  Score 14.4;  DB 3;  Length 4342;
  Best Local Similarity   93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          ||||||| |||||||||
Db   1503 GGGGCCGGCCTGGTGC 1488


RESULT   7
PCT-US95-08812-1/c
; Sequence 1, Application PC/TUS9508812
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  HTK LIGAND
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/08812
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  902PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4342 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US95-08812-1


  Query Match             84.7%;  Score 14.4;  DB 6;  Length 4342;
  Best Local Similarity   93.8%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           ||||||| ||||||||
Db    1503 GGGGCCGGCCTGGTGC 1488


RESULT   8
US-07-807-043B-5/c
; Sequence 5, Application US/07807043B
; Patent No. 5342774
;  GENERAL INFORMATION:
;    APPLICANT:  Boon, Thierry, Van den Eynde, Beno t
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursors, Tumor
;    TITLE OF INVENTION:  Rejection Antigens and Uses Thereof
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/807,043B
;      FILING DATE:  19911212
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5342774man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 253.3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4698 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  singular
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-07-807-043B-5


  Query Match             84.7%;  Score 14.4;  DB 1;  Length 4698;
  Best Local Similarity   93.8%;  Pred. No. 97;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           ||||| ||||||||||
Db    1709 GGGGCAGTCCTGGTGC 1694


RESULT   9
US-08-299-849B-5/c
; Sequence 5, Application US/08299849B
; Patent No. 5612201
;  GENERAL INFORMATION:
;    APPLICANT:  De Plaen, Etienne; Boon-Falleur, Thierry;
;    APPLICANT:  Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;    APPLICANT:  Chomez, Patrick
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Useful In
;    TITLE OF INVENTION:  Determining Expression Of A Tumor Antigen Precursor
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/299,849B
;      FILING DATE:  1-SEPTEMBE  1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5612201man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5355
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4698 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  genomic DNA
US-08-299-849B-5


  Query Match             84.7%;  Score 14.4;  DB 1;  Length 4698;
  Best Local Similarity   93.8%;  Pred. No. 97;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           ||||| ||||||||||
Db    1709 GGGGCAGTCCTGGTGC 1694


RESULT  10
US-08-142-368A-5/c
; Sequence 5, Application US/08142368A
; Patent No. 5925729
;  GENERAL INFORMATION:
;    APPLICANT:  Boon-Falleur, Thierry; Van der Bruggen, Thierry;
;    APPLICANT:  Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
;    APPLICANT:  Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursors, Tumor
;    TITLE OF INVENTION:  Rejection Antigens and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/142,368A
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5925729man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5253.4-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4698 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-142-368A-5


  Query Match             84.7%;  Score 14.4;  DB 4;  Length 4698;
  Best Local Similarity   93.8%;  Pred. No. 97;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           ||||| ||||||||||
Db    1709 GGGGCAGTCCTGGTGC 1694


RESULT  11
US-08-967-727-5/c
; Sequence 5, Application US/08967727
; Patent No. 6025474
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,727
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,365
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6025474man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4698 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-967-727-5


  Query Match             84.7%;  Score 14.4;  DB 5;  Length 4698;
  Best Local Similarity   93.8%;  Pred. No. 97;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           ||||| ||||||||||
Db    1709 GGGGCAGTCCTGGTGC 1694
```

```
RESULT 12
US-08-932-280-1/c
; Sequence 1, Application US/08932280
; Patent No. 6013486
; GENERAL INFORMATION:
;   APPLICANT: Dellaporta, Stephen L.
;   TITLE OF INVENTION: METHOD OF SELECTION OF INSERTION
;   TITLE OF INVENTION: MUTATIONS
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Arnold, White & Durkee
;     STREET: P.O. Box 4433
;     CITY: Houston
;     STATE: TX
;     COUNTRY: US
;     ZIP: 77210-4433
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/932,280
;     FILING DATE: CONCURRENTLY HEREWITH
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Highlander, Steven L.
;     REGISTRATION NUMBER: 37,642
;     REFERENCE/DOCKET NUMBER: DEKM:156
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (512) 418-3000
;     TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 6743 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-08-932-280-1


  Query Match            82.4%;  Score 14;  DB 5;  Length 6743;
  Best Local Similarity  100.0%;  Pred. No. 1.4e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 GGCCGTCCTGGTGC 16
           ||||||||||||||
Db     906 GGCCGTCCTGGTGC 893


RESULT 13
US-08-105-168B-1/c
; Sequence 1, Application US/08105168B
; Patent No. 5589585
; GENERAL INFORMATION:
;   APPLICANT: MABILAT et al.
;   TITLE OF INVENTION: DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;   TITLE OF INVENTION: PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;   TITLE OF INVENTION: MYCOBACTERIA
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Oliff & Berridge
;     STREET: 700 South Washington Street, Suite 300
;     CITY: Alexandria,
;     STATE: Virginia
;     ZIP: 22314
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5" DS/HD
;     COMPUTER: IBM compatible
;     OPERATING SYSTEM: MS DOS 3.1
;     SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;     FILING DATE: August 12, 1993
;     APPLICATION NUMBER: US/08/105,168B
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: FR9210094
;     FILING DATE: August 8, 1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: William P. Berridge
;     REGISTRATION NUMBER: 30,024
;     REFERENCE/DOCKET NUMBER: WPB 28835
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 836-6400
;     TELEFAX: (703) 836-2787
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 268 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single-stranded
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   ORIGINAL SOURCE:
;     ORGANISM: Mycobacterium bovis
;     STRAIN:
;     INDIVIDUAL ISOLATE:
;   POSITION IN GENOME:
;     CHROMOSOME/SEGMENT:
;     MAP POSITION:
;   FEATURE:
;     NAME/KEY:
;     LOCATION: 461-728
;     IDENTIFICATION METHOD:
;     OTHER INFORMATION:
US-08-105-168B-1


  Query Match            81.2%;  Score 13.8;  DB 1;  Length 268;
  Best Local Similarity  88.2%;  Pred. No. 2.2e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           | ||||||| |||||||
Db      66 GTGGCCGTCGTGGTGCC 50


RESULT 14
US-08-105-168B-2/c
; Sequence 2, Application US/08105168B
; Patent No. 5589585
; GENERAL INFORMATION:
;   APPLICANT: MABILAT et al.
;   TITLE OF INVENTION: DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;   TITLE OF INVENTION: PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE
;   TITLE OF INVENTION: MYCOBACTERIA
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Oliff & Berridge
;     STREET: 700 South Washington Street, Suite 300
;     CITY: Alexandria,
;     STATE: Virginia
;     ZIP: 22314
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5" DS/HD
;     COMPUTER: IBM compatible
;     OPERATING SYSTEM: MS DOS 3.1
;     SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;     FILING DATE: August 12, 1993
;     APPLICATION NUMBER: US/08/105,168B
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  FR9210094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 28835
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium microti
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  461-728
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-105-168B-2


  Query Match           81.2%;  Score 13.8;  DB 1;  Length 268;
  Best Local Similarity  88.2%;  Pred. No. 2.2e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           | ||||||| |||||||
Db      66 GTGGCCGTCGTGGTGCC 50


RESULT  15
US-08-105-168B-3/c
; Sequence 3, Application US/08105168B
; Patent No. 5589585
;  GENERAL INFORMATION:
;    APPLICANT:  MABILAT et al.
;    TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;    TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;    TITLE OF INVENTION:  MYCOBACTERIA
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oliff & Berridge
;      STREET:  700 South Washington Street, Suite 300
;      CITY:  Alexandria,
;      STATE:  Virginia
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" DS/HD
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS DOS 3.1
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      FILING DATE:  August 12, 1993
;      APPLICATION NUMBER:  US/08/105,168B
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR9210094
;      FILING DATE:  August 8, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William P. Berridge
;      REGISTRATION NUMBER:  30,024
;      REFERENCE/DOCKET NUMBER:  WPB 28835
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6400
;      TELEFAX:  (703) 836-2787
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycobacterium africanum
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:  461-728
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-105-168B-3


  Query Match           81.2%;  Score 13.8;  DB 1;  Length 268;
  Best Local Similarity  88.2%;  Pred. No. 2.2e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           | ||||||| |||||||
Db      66 GTGGCCGTCGTGGTGCC 50


Search completed: June 22, 2000, 15:22:00
Job time: 18798 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:20 ; Search time 5541.94 Seconds
(without alignments)
18.284 Million cell updates/sec

Title: US-09-362-485-14
Perfect score: 25
Sequence: 1 GAATTCCCATCAGCAATCTTGCAGA 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	22.4	89.6	504	109	AQ613871	AQ613871 HS_5136_B
	2	20.2	80.8	812	48	AI554378	AI554378 tq16f12.x
c	3	19.8	79.2	390	50	AU066317	AU066317 AU066317
c	4	19.8	79.2	400	20	D40152	D40152 RICS1926A R
c	5	19.8	79.2	424	50	AU066205	AU066205 AU066205
c	6	19.8	79.2	437	43	AU033304	AU033304 AU033304
c	7	19.8	79.2	439	20	D41627	D41627 RICS4255A R
c	8	19.8	79.2	445	20	D41513	D41513 RICS4052A R
c	9	19.8	79.2	450	20	D41157	D41157 RICS3462A R
c	10	19.8	79.2	460	34	C25324	C25324 C25324 Rice
c	11	19.8	79.2	490	51	AU070635	AU070635 AU070635
c	12	19.4	77.6	502	51	AI759818	AI759818 sb64f07.y
c	13	19.2	76.8	244	41	AI036833	AI036833 vw96c08.r
c	14	19.2	76.8	351	69	AW127023	AW127023 ga20a09.y
c	15	19.2	76.8	698	107	AQ484028	AQ484028 RPCI-11-2
c	16	18.8	75.2	406	20	D41156	D41156 RICS3459A R
	17	18.6	74.4	198	73	AV342685	AV342685 AV342685
	18	18.6	74.4	254	74	AV380702	AV380702 AV380702
	19	18.6	74.4	289	70	AV252621	AV252621 AV252621
	20	18.6	74.4	294	72	AV307847	AV307847 AV307847
	21	18.6	74.4	296	72	AV307979	AV307979 AV307979
	22	18.6	74.4	299	72	AV300061	AV300061 AV300061
	23	18.6	74.4	431	91	AQ868222	AQ868222 nbeb0019P
	24	18.6	74.4	479	91	AQ870483	AQ870483 nbeb0039I
	25	18.6	74.4	811	91	AQ868470	AQ868470 nbeb0027J
	26	18.6	74.4	1060	91	AQ900249	AQ900249 HS_2054_B
c	27	18.4	73.6	468	25	W01577	W01577 za41a03.r1
c	28	18.4	73.6	538	108	AQ515203	AQ515203 HS_2208_B
c	29	18.2	72.8	297	20	D40212	D40212 RICS2021A R
c	30	18.2	72.8	309	20	D40893	D40893 RICS3075A R
c	31	18.2	72.8	312	20	D40897	D40897 RICS3081A R
c	32	18.2	72.8	323	20	D40245	D40245 RICS2079A R
c	33	18.2	72.8	324	20	D40679	D40679 RICS2787A R
c	34	18.2	72.8	342	50	AU066204	AU066204 AU066204
c	35	18.2	72.8	344	20	D41544	D41544 RICS4093A R
c	36	18.2	72.8	350	25	N88367	N88367 K3176F Huma
c	37	18.2	72.8	351	20	D41352	D41352 RICS3824A R
c	38	18.2	72.8	365	20	D39662	D39662 RICS1215A R
c	39	18.2	72.8	369	20	D41373	D41373 RICS3850A R
c	40	18.2	72.8	382	20	D41302	D41302 RICS3692A R
c	41	18.2	72.8	384	50	AU066212	AU066212 AU066212
c	42	18.2	72.8	388	20	D40557	D40557 RICS2619A R
c	43	18.2	72.8	389	50	AU066209	AU066209 AU066209
c	44	18.2	72.8	389	50	AU066316	AU066316 AU066316
c	45	18.2	72.8	395	50	AU066133	AU066133 AU066133

ALIGNMENTS

RESULT 1
AQ613871
LOCUS AQ613871 504 bp DNA GSS 15-JUN-1999
DEFINITION HS_5136_B1_D02_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=712 Col=3 Row=H, genomic survey sequence.
ACCESSION AQ613871
VERSION AQ613871.1 GI:5075147
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 504)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 712 row: H column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 504.
FEATURES Location/Qualifiers
source 1. .504
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=712 Col=3 Row=H"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 141 a 104 c 95 g 161 t 3 others
ORIGIN

Query Match 89.6%; Score 22.4; DB 109; Length 504;
Best Local Similarity 95.8%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAATTCC ATCAGCAATCTTGCAG 24
|||| || ||||||||||||||||
Db 445 GAATCCC ATCAGCAATCTTGCAG 468

RESULT 2
AI554378
LOCUS AI554378 812 bp mRNA EST 13-MAY-1999
DEFINITION tq16f12.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2209007 3' similar to gb:M73700 LACTOTRANSFERRIN PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION AI554378
VERSION AI554378.1 GI:4486741
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 812)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 7, 1998 this sequence version replaced gi:3121298.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2212 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 397
POLYA=No.
FEATURES Location/Qualifiers

```
     source          1. .812
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2209007"
                     /clone_lib="NCI_CGAP_Ut3"
                     /tissue_type="poorly-differentiated endometrial
                     adenocarcinoma, 2 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.45 kb. Life Technologies catalog #:
                     11541-018"
BASE COUNT      201 a    202 c    203 g    202 t      4 others
ORIGIN


  Query Match             80.8%;  Score 20.2;  DB 48;  Length 812;
  Best Local Similarity   88.0%;  Pred. No. 17;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGA 25
          | ||||||||||||  |||||||||
Db    597 GTATTCCCATCAGCGTTCTTGCAGA 621


RESULT   3
AU066317/c
LOCUS       AU066317      390 bp    mRNA            EST       02-JUN-1999
DEFINITION  AU066317 Rice shoot Oryza sativa cDNA clone S6424_1A, mRNA
            sequence.
ACCESSION   AU066317
VERSION     AU066317.1  GI:4969413
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138655.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S6424_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT      101 a     91 c     92 g    103 t      3 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 50;  Length 390;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AATTCCCATCAGCAATCTTGCAG 24
          |||||||||| |||| ||||||
Db    265 AATTCCCATCAACAATTTTGCAG 243


RESULT   4
D40152/c
LOCUS       D40152        400 bp    mRNA            EST       11-NOV-1994
DEFINITION  RICS1926A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D40152
VERSION     D40152.1  GI:569303
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT      106 a     92 c     96 g    105 t      1 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 20;  Length 400;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AATTCCCATCAGCAATCTTGCAG 24
          |||||||||| |||| ||||||
Db    275 AATTCCCATCAACAATTTTGCAG 253


RESULT   5
AU066205/c
LOCUS       AU066205      424 bp    mRNA            EST       02-JUN-1999
DEFINITION  AU066205 Rice shoot Oryza sativa cDNA clone S5610_1A, mRNA
            sequence.
ACCESSION   AU066205
VERSION     AU066205.1  GI:4969301
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138543.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
```

```
           PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S5610_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT      111 a     97 c    102 g    111 t      3 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 50;  Length 424;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AATTCCCATCAGCAATCTTGCAG 24
          ||||||||||| |||| ||||||
Db    266 AATTCCCATCAACAATTTTGCAG 244


RESULT   6
AU033304/c
LOCUS       AU033304       437 bp    mRNA            EST       20-OCT-1998
DEFINITION  AU033304 Rice shoot Oryza sativa cDNA clone S6503_1A, mRNA
            sequence.
ACCESSION   AU033304
VERSION     AU033304.1  GI:3769277
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900405.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S6503_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT      115 a     103 c    103 g    113 t      3 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 43;  Length 437;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AATTCCCATCAGCAATCTTGCAG 24
          ||||||||||| |||| ||||||
Db    267 AATTCCCATCAACAATTTTGCAG 245


RESULT   7
D41627/c
LOCUS       D41627         439 bp    mRNA            EST       15-NOV-1994
DEFINITION  RICS4255A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D41627
VERSION     D41627.1  GI:572116
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT      112 a     107 c    106 g    112 t      2 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 20;  Length 439;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  ndels    0;  Gaps    0;

Qy      2 AATTCCCATCAGCAATCTTGCAG 24
          ||||||||||| |||| ||||||
Db    260 AATTCCCATCAACAATTTTGCAG 238


RESULT   8
D41513/c
LOCUS       D41513         445 bp    mRNA            EST       15-NOV-1994
DEFINITION  RICS4052A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D41513
VERSION     D41513.1  GI:572002
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
```

```
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       117 a    104 c    110 g    114 t
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 20;  Length 445;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAG 24
           ||||||||||| |||| ||||||
Db     279 AATTCCCATCAACAATTTTGCAG 257


RESULT   9
D41157/c
LOCUS       D41157         450 bp    mRNA            EST       15-NOV-1994
DEFINITION  RICS3462A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D41157
VERSION     D41157.1  GI:571646
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       119 a    104 c    108 g    116 t      3 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 20;  Length 450;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAG 24
           ||||||||||| |||| ||||||
Db     260 AATTCCCATCAACAATTTTGCAG 238


RESULT  10
C25324/c
LOCUS       C25324         460 bp    mRNA            EST       25-JUL-1997
DEFINITION  C25324 Rice shoot Oryza sativa cDNA clone S5942_1A, mRNA sequence.
ACCESSION   C25324
VERSION     C25324.1  GI:2280829
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637865.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S5942_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       121 a    107 c    111 g    121 t
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 34;  Length 460;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAG 24
           ||||||||||| |||| ||||||
Db     268 AATTCCCATCAACAATTTTGCAG 246


RESULT  11
AU070635/c
LOCUS       AU070635       490 bp    mRNA            EST       07-JUN-1999
DEFINITION  AU070635 Rice shoot Oryza sativa cDNA clone S6271_3A, mRNA
            sequence.
ACCESSION   AU070635
VERSION     AU070635.1  GI:5005473
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189024.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S6271_3A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       129 a    115 c    118 g    127 t      1 others
ORIGIN
```

```
  Query Match             79.2%;  Score 19.8;  DB 51;  Length 490;
  Best Local Similarity   91.3%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAG 24
           |||||||||||  |||| ||||||
Db     265 AATTCCCATCAACAATTTTGCAG 243


RESULT  12
AI759818/c
LOCUS       AI759818      502 bp    mRNA            EST       24-JUN-1999
DEFINITION  sb64f07.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1017-158 5', mRNA sequence.
ACCESSION   AI759818
VERSION     AI759818.1  GI:5175485
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On May 1, 1997 this sequence version replaced gi:2059646.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-158"
                     /clone_lib="Gm-c1017"
                     /tissue_type="vegetable buds of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from vegatable buds of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      149 a     93 c    126 g    134 t
ORIGIN


  Query Match             77.6%;  Score 19.4;  DB 51;  Length 502;
  Best Local Similarity   95.2%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 TTCCCATCAGCAATCTTGCAG 24
           || ||||||||||||||||||
Db     403 TTTCCATCAGCAATCTTGCAG 383


RESULT  13
AI036833/c
LOCUS       AI036833      244 bp    mRNA            EST       26-JUN-1998
DEFINITION  vw96c08.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:1262798 5' similar to TR:Q62220 Q62220 SERINE 2 ULTRA HIGH
            SULFUR PROTEIN. ;, mRNA sequence.
ACCESSION   AI036833
VERSION     AI036833.1  GI:3260524
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Dec 18, 1997 this sequence version replaced gi:2339790.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:665350
            Seq primer: -28m13 revl ET from Amersham
            High quality sequence stop: 215.
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1262798"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       35 a     86 c     61 g     62 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 41;  Length 244;
  Best Local Similarity   87.5%;  Pred. No. 42;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           |||||||||  |||||  |||||||
Db     109 GAATTCCCAGCAGCAGGCTTGCAG 86


RESULT  14
```

```
AW127023/c
LOCUS       AW127023      351 bp    mRNA            EST       22-OCT-1999
DEFINITION  ga20a09.y1 Moss EST library PPU Physcomitrella patens cDNA clone
            PEP_SOURCE_ID:PPU021317 5', mRNA sequence.
ACCESSION   AW127023
VERSION     AW127023.1  GI:6102553
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
  ORGANISM  Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C.,
            Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T.,
            Underwood,K., Theising,B., Allen,M., Bowers,Y.I, Person,B.,
            Swaller,T., Steptce,M., Gibbons,M., Harvey,N., Ritter,E.,
            Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
  TITLE     Leeds/Wash U Moss EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137159.
            Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashiardes as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 346.
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Physcomitrella patens"
                     /db_xref="taxon:3218"
                     /clone="PEP_SOURCE_ID:PPU021317"
                     /clone_lib="Moss EST library PPU"
                     /tissue_type="protonemata: 7 day old tissue
                     ammonium-grown"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Construction of the cDNA library was carried out
                     using Stratagenes 'UniZAP - cDNA synthesis kit'. cDNA
                     was constructed using an oligo dT primer/linker that
                     contains a XhoI site within it. Following ds cDNA
                     synthesis, EcoRI adapters were ligated to the blunt ends
                     and sample was digested with XhoI. The result is cDNA
                     with an EcoRI sticky end on one side and a XhoI sticky
                     end on the other. This cDNA was ligated directionally in
                     UniZAP arms. The vector is designed containing the
                     pBluescript sequence as well as lambda DNA and cDNA is
                     cloned within this pBluescript sequence. The vector was
                     then packaged using Gold gigapackaging extracts. Library
                     was grown in XLIBlue MRF' cells and amplified. The library
                     was excised by mass excision using Stratagens 'Mass
                     excision kit' that uses exassist as a helper phage that
                     releases the pBluescript sequence and circularises it as
                     single stranded plasmids that are then packaged (by helper
                     phage) and secreted out of the host cell as phagemids.
                     SOLR cells were transformed with phagemids and the library
                     was plated out on LB-amp plates to select for
                     transformants. Approximately 1,000,000 colonies were grown
                     and recovered. The double stranded plasmid library was
                     recovered by using Quiagen Midi prep kit. 2 micro grams of
                     each library were used to transform DH10B cells by
                     electroporation."
BASE COUNT       83 a    112 c     81 g     75 t
ORIGIN
```

```
  Query Match           76.8%;  Score 19.2;  DB 69;  Length 351;
  Best Local Similarity  87.5%;  Pred. No. 44;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           ||| |||| ||||||||||| ||
Db     329 GAAATCCCCTCAGCAATCTTGAAG 306


RESULT  15
AQ484028/c
LOCUS       AQ484028      698 bp    DNA             GSS       24-APR-1999
DEFINITION  RPCI-11-265F13.TJ RPCI-11 Homo sapiens genomic clone
            RPCI-11-265F13, genomic survey sequence.
ACCESSION   AQ484028
VERSION     AQ484028.1  GI:4671349
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-265F13.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .698
                     /organism="Homo sapiens"
                     /db_xref="GDB:7601508"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-265F13"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      145 a    133 c     96 g    324 t
ORIGIN
```

```
  Query Match           76.8%;  Score 19.2;  DB 107;  Length 698;
  Best Local Similarity  87.5%;  Pred. No. 49;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           || |||||| ||| ||||||||||
Db     186 GACTTCCCAGCAGAAATCTTGCAG 163


Search completed: June 22, 2000, 12:08:22
Job time: 10416 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:31 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-15
Perfect score: 18
Sequence: 1 GCCCCGATGAGCGAAGTC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87618	A87618 Sequence 15
2		18	100.0	720	5	A89759	A89759 Sequence 16
3		18	100.0	1194	2	MTU92472	U92472 Mycobacteri
4		18	100.0	1208	5	A89752	A89752 Sequence 9
5		18	100.0	1209	5	A87612	A87612 Sequence 9
6		18	100.0	1228	5	A87608	A87608 Sequence 5
7		18	100.0	1228	5	A89748	A89748 Sequence 5
8		18	100.0	1229	5	A87610	A87610 Sequence 7
9		18	100.0	1229	5	A89750	A89750 Sequence 7
10		18	100.0	1235	5	A87606	A87606 Sequence 3
11		18	100.0	1235	5	A87609	A87609 Sequence 6
12		18	100.0	1235	5	A87611	A87611 Sequence 8
13		18	100.0	1235	5	A89746	A89746 Sequence 3
14		18	100.0	1235	5	A89749	A89749 Sequence 6
15		18	100.0	1235	5	A89751	A89751 Sequence 8
16		18	100.0	1236	5	A87613	A87613 Sequence 10
17		18	100.0	1236	5	A89753	A89753 Sequence 10
18		18	100.0	1237	5	A87607	A87607 Sequence 4
19		18	100.0	1237	5	A89747	A89747 Sequence 4
20		18	100.0	1245	5	A87605	A87605 Sequence 2
21		18	100.0	1245	5	A89745	A89745 Sequence 2
22		18	100.0	1260	5	A87604	A87604 Sequence 1
23		18	100.0	1260	5	A89744	A89744 Sequence 1
24		18	100.0	2584	1	MTALADH	X63069 M.tuberculo
25		18	100.0	56414	1	MTV002	AL008967 Mycobacte
26	c	15.4	85.6	2700	2	AF152394	AF152394 Mycobacte
27		15	83.3	20770	34	DRORDR	D28563 Fruit fly D
28		15	83.3	65114	34	AB011813	AB011813 Drosophil
29	c	15	83.3	82818	42	AC012754	AC012754 Drosophil
30	c	15	83.3	151373	42	AC008308	AC008308 Drosophil
31	c	14.8	82.2	460	3	AF055516	AF055516 Didelphis
32	c	14.8	82.2	509	8	AF070019	AF070019 Lecanora
33		14.8	82.2	682	5	A89758	A89758 Sequence 15
34		14.8	82.2	682	5	A89762	A89762 Sequence 19
35	c	14.8	82.2	1174	40	AF118452	AF118452 Homo sapi
36		14.8	82.2	1941	4	APHISH1	X06128 Duck histon
37	c	14.8	82.2	2142	11	HSU31468	U31468 Homo sapien
38	c	14.8	82.2	4624	10	HSL3G9A	Z69387 Human DNA s
39	c	14.8	82.2	5482	2	BCU19883	U19883 Burkholderi
40		14.8	82.2	5596	1	NPSDHCDBA	Y07709 N.pharaonis
41	c	14.8	82.2	7011	5	I96205	I96205 Sequence 42
42	c	14.8	82.2	7011	9	HUMMYC3L	M19720 Human L-myc
43	c	14.8	82.2	10401	2	AE001934	AE001934 Deinococc
44	c	14.8	82.2	29655	44	AC020348	AC020348 Drosophil
45		14.8	82.2	38084	1	SCE68	AL079345 Streptomy

ALIGNMENTS

```
RESULT   1
A87618
LOCUS       A87618         18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 15 from Patent WO9836089.
ACCESSION   A87618
VERSION     A87618.1  GI:6736258
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      6 c      6 g      2 t
ORIGIN

  Query Match           100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           ||||||||||||||||||
Db       1 GCCCCGATGAGCGAAGTC 18


RESULT   2
A89759
LOCUS       A89759         720 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9832862.
ACCESSION   A89759
VERSION     A89759.1  GI:6738291
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      133 a    238 c    226 g    123 t
ORIGIN

  Query Match           100.0%;  Score 18;  DB 5;  Length 720;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           ||||||||||||||||||
Db     384 GCCCCGATGAGCGAAGTC 401


RESULT   3
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN

  Query Match           100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           ||||||||||||||||||
Db     451 GCCCCGATGAGCGAAGTC 468


RESULT   4
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT   5
A87612
LOCUS       A87612       1209 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT   6
A87608
LOCUS       A87608       1228 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT   7
A89748
LOCUS       A89748       1228 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT   8
A87610
LOCUS       A87610       1229 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467


RESULT   9
A89750
LOCUS       A89750       1229 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
```

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467


RESULT  10
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT  11
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467


RESULT  12
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467


RESULT  13
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

```
  Query Match           100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    451 GCCCCGATGAGCGAAGTC 468


RESULT  14
A89749
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match           100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467


RESULT  15
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match           100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||||||||||||||||
Db    450 GCCCCGATGAGCGAAGTC 467
```

Search completed: June 22, 2000, 14:58:33
Job time: 17928 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:38 ; Search time 341.15 Seconds
(without alignments)
9.526 Million cell updates/sec

Title: US-09-362-485-14
Perfect score: 25
Sequence: 1 GAATTCCCATCAGCAATCTTGCAGA 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16.6	66.4	807	2	US-08-270-584A-1	Sequence 1, Appli
	2	16.6	66.4	807	3	US-08-765-192-1	Sequence 1, Appli
	3	16.6	66.4	959	2	US-08-568-031-1	Sequence 1, Appli
	4	16.6	66.4	959	3	US-08-966-319-1	Sequence 1, Appli
c	5	16.6	66.4	2188	1	US-07-882-925A-4	Sequence 4, Appli
c	6	16.6	66.4	2188	1	US-08-184-012C-4	Sequence 4, Appli
c	7	16.6	66.4	2266	3	US-08-724-394A-18	Sequence 18, Appl
c	8	16.6	66.4	6751	1	US-07-882-925A-5	Sequence 5, Appli
c	9	16.6	66.4	6751	1	US-08-184-012C-5	Sequence 5, Appli
	10	16.6	66.4	246240	3	US-08-724-394A-20	Sequence 20, Appl
	11	16.6	66.4	246240	3	US-08-724-394A-21	Sequence 21, Appl
	12	16.6	66.4	246240	3	US-08-724-394A-22	Sequence 22, Appl
c	13	16.2	64.8	1325	1	US-08-547-182-1	Sequence 1, Appli
c	14	16	64.0	46	1	US-08-427-640-28	Sequence 28, Appl
c	15	16	64.0	862	5	US-08-713-569-4	Sequence 4, Appli
c	16	16	64.0	1068	1	US-08-427-640-7	Sequence 7, Appli
c	17	16	64.0	2200	2	US-08-272-255-21	Sequence 21, Appl
c	18	16	64.0	2200	6	PCT-US95-08565-21	Sequence 21, Appl
c	19	15.8	63.2	2003	6	PCT-US95-07085-1	Sequence 1, Appli
	20	15.6	62.4	2012	4	US-08-555-568B-16	Sequence 16, Appl
	21	15.6	62.4	2109	4	US-08-555-568B-20	Sequence 20, Appl
	22	15.6	62.4	2112	4	US-08-555-568B-22	Sequence 22, Appl
c	23	15.6	62.4	3889	3	US-08-648-298-1	Sequence 1, Appli
c	24	15.6	62.4	4875	2	US-08-460-739-1	Sequence 1, Appli
	25	15.4	61.6	324	4	US-08-853-659A-34	Sequence 34, Appl
	26	15.4	61.6	921	2	US-08-420-235B-30	Sequence 30, Appl
	27	15.4	61.6	921	6	PCT-US95-10194-30	Sequence 30, Appl
c	28	15.4	61.6	1581	4	US-08-853-659A-15	Sequence 15, Appl
c	29	15.4	61.6	1785	4	US-08-559-505-3	Sequence 3, Appli
c	30	15.4	61.6	1785	4	US-08-749-907-3	Sequence 3, Appli
c	31	15.4	61.6	1785	6	PCT-US95-07721-7	Sequence 7, Appli
c	32	15.4	61.6	1939	7	5198542-3	Patent No. 5198542
c	33	15.4	61.6	7876	2	US-08-225-480-1	Sequence 1, Appli
c	34	15.4	61.6	7876	6	PCT-US93-08115-1	Sequence 1, Appli
c	35	15.4	61.6	15512	4	US-08-853-659A-5	Sequence 5, Appli
	36	15.4	61.6	15512	4	US-08-853-659A-8	Sequence 8, Appli
c	37	15.4	61.6	15512	4	US-08-853-659A-63	Sequence 63, Appl
	38	15.4	61.6	15512	4	US-08-853-659A-66	Sequence 66, Appl
c	39	15.4	61.6	17041	1	US-08-076-011-1	Sequence 1, Appli
c	40	15.4	61.6	20710	2	US-08-420-235B-1	Sequence 1, Appli
c	41	15.4	61.6	20710	6	PCT-US95-10194-1	Sequence 1, Appli
c	42	15.4	61.6	24701	4	US-08-853-659A-2	Sequence 2, Appli
	43	15.4	61.6	24701	4	US-08-853-659A-3	Sequence 3, Appli
c	44	15.4	61.6	24701	4	US-08-853-659A-60	Sequence 60, Appl
	45	15.4	61.6	24701	4	US-08-853-659A-61	Sequence 61, Appl

ALIGNMENTS

```
RESULT   1
US-08-270-584A-1
; Sequence 1, Application US/08270584A
; Patent No. 5710035
;  GENERAL INFORMATION:
;    APPLICANT:  GREENE, ET AL.
;    TITLE OF INVENTION:  Human Elastase IV
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/270,584A
;      FILING DATE:  July 5, 1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-126
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA
US-08-270-584A-1

  Query Match              66.4%;  Score 16.6;  DB 2;  Length 807;
  Best Local Similarity    82.6%;  Pred. No. 22;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
```

```
          ||| || ||| ||| ||||||||
Db    364 ATTGCCCTCATCAAGCTTGCAGA 386


RESULT   2
US-08-765-192-1
; Sequence 1, Application US/08765192
; Patent No. 5851814
;  GENERAL INFORMATION:
;    APPLICANT:  Greene, John et al.
;    TITLE OF INVENTION:  Human Elastase IV
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,192
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF125US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3013098504
;      TELEFAX:  3013098439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-765-192-1


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 807;
  Best Local Similarity   82.6%;  Pred. No. 22;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     364 ATTGCCCTCATCAAGCTTGCAGA 386


RESULT   3
US-08-568-031-1
; Sequence 1, Application US/08568031
; Patent No. 5738991
;  GENERAL INFORMATION:
;    APPLICANT:  Braxton, Scott M.
;    APPLICANT:  Diep, Dinh
;    APPLICANT:  Delegeane, Angelo M.
;    TITLE OF INVENTION:  HOMOLOG OF RAT ELASTASE IV DERIVED FROM
;    TITLE OF INVENTION:  HUMAN PANCREAS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3330 Hillview Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/568,031
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luther, Barbara J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PF-0046 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-852-0195
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  959 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      LIBRARY:  HUMAN ELASTASE HOMOLOG
;      CLONE:  226990
US-08-568-031-1


  Query Match             66.4%;  Score 16.6;  DB 2;  Length 959;
  Best Local Similarity   82.6%;  Pred. No. 22;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     385 ATTGCCCTCATCAAGCTTGCAGA 407


RESULT   4
US-08-966-319-1
; Sequence 1, Application US/08966319
; Patent No. 5856109
;  GENERAL INFORMATION:
;    APPLICANT:  Braxton, Scott M.
;    APPLICANT:  Diep, Dinh
;    APPLICANT:  Delegeane, Angelo M.
;    TITLE OF INVENTION:  HOMOLOG OF RAT ELASTASE IV DERIVED FROM
;    TITLE OF INVENTION:  HUMAN PANCREAS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3330 Hillview Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,319
;      FILING DATE:  07-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/568,031
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luther, Barbara J.
;      REGISTRATION NUMBER:  33954
;      REFERENCE/DOCKET NUMBER:  PF-0046 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
```

```
;      TELEFAX:  415-852-0195
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  959 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       LIBRARY:  HUMAN ELASTASE HOMOLOG
;       CLONE:  226990
US-08-966-319-1


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 959;
  Best Local Similarity   82.6%;  Pred. No. 22;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     385 ATTGCCCTCATCAAGCTTGCAGA 407


RESULT   5
US-07-882-925A-4/c
; Sequence 4, Application US/07882925A
; Patent No. 5315000
;  GENERAL INFORMATION:
;    APPLICANT:  Degen, Sandra J. F.
;    TITLE OF INVENTION:  Gene for a growth factor and its cDNA and
;    TITLE OF INVENTION:  protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gregory Lunn
;      STREET:  Wood, Herron & Evans, 2700 Carew Tower
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,  800 Kb
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.3
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/882,925A
;      FILING DATE:  19920514
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Gregory
;      REGISTRATION NUMBER:  29,945
;      REFERENCE/DOCKET NUMBER:  CMC 57
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 241-2324
;      TELEFAX:  (513) 421-7269
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2188 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;      STRAIN:  C57BL/6
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  liver
;    IMMEDIATE SOURCE:
;      LIBRARY:  cDNA
;      CLONE:  ML5-2
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  mouse 9, Hgfl locus
;      MAP POSITION:  Trf-Gnai-2-Hgfl-Cck
;    FEATURE:
;      IDENTIFICATION METHOD:  experimental
;    PUBLICATION INFORMATION:
;      RELEVANT RESIDUES IN SEQ ID NO:  4: 1 TO 2188
US-07-882-925A-4


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 2188;
  Best Local Similarity   82.6%;  Pred. No. 26;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           || |||||||| | || ||||||
Db    1258 ATCCCCATCAGGATTCCTGCAGA 1236


RESULT   6
US-08-184-012C-4/c
; Sequence 4, Application US/08184012C
; Patent No. 5606029
;  GENERAL INFORMATION:
;    APPLICANT:  Degen, Sandra J. F.
;    TITLE OF INVENTION:  Gene for a growth factor and its cDNA and
;    TITLE OF INVENTION:  protein
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gregory Lunn
;      STREET:  Wood, Herron & Evans, 2700 Carew Tower
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,  800 Kb
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5.2
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,012C
;      FILING DATE:  1/18/94
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Gregory
;      REGISTRATION NUMBER:  29,945
;      REFERENCE/DOCKET NUMBER:  CMC 57
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 241-2324
;      TELEFAX:  (513) 421-7269
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2188 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;      STRAIN:  C57BL/6
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  liver
;    IMMEDIATE SOURCE:
;      LIBRARY:  cDNA
;      CLONE:  ML5-2
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  mouse 9, Hgfl locus
;      MAP POSITION:  Trf-Gnai-2-Hgfl-Cck
;    FEATURE:
;      IDENTIFICATION METHOD:  experimental
;    PUBLICATION INFORMATION:
;      RELEVANT RESIDUES IN SEQ ID NO:  4: 1 TO 2188
US-08-184-012C-4
```

```
  Query Match             66.4%;  Score 16.6;  DB 1;  Length 2188;
  Best Local Similarity   82.6%;  Pred. No. 26;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           || |||||||| | || ||||||
Db    1258 ATCCCCATCAGGATTCCTGCAGA 1236


RESULT   7
US-08-724-394A-18/c
; Sequence 18, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2266 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2266
;      OTHER INFORMATION:  /note= "cDNA 22B"
US-08-724-394A-18


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 2266;
  Best Local Similarity   82.6%;  Pred. No. 26;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           |||||||| || || |||||| |
Db    1505 ATTCCCATGAGGAAACTTGCATA 1483


RESULT   8
```

```
US-07-882-925A-5/c
; Sequence 5, Application US/07882925A
; Patent No. 5315000
;  GENERAL INFORMATION:
;    APPLICANT:  Degen, Sandra J. F.
;    TITLE OF INVENTION:  Gene for a growth factor and its cDNA and
;    TITLE OF INVENTION:  protein
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gregory Lunn
;      STREET:  Wood, Herron & Evans, 2700 Carew Tower
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,  800 Kb
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.3
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/882,925A
;      FILING DATE:  19920514
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Gregory
;      REGISTRATION NUMBER:  29,945
;      REFERENCE/DOCKET NUMBER:  CMC 57
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 241-2324
;      TELEFAX:  (513) 421-7269
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6751 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;      STRAIN:  Balb/c
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  liver
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic
;      CLONE:  MGL5-12
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  mouse 9, Hgfl locus
;      MAP POSITION:  Trf-Gnai-2-Hgfl-Cck
;    FEATURE:
;      IDENTIFICATION METHOD:  experimental
;    PUBLICATION INFORMATION:
;      RELEVANT RESIDUES IN SEQ ID NO:  5: 1 TO 6751
US-07-882-925A-5


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 6751;
  Best Local Similarity   82.6%;  Pred. No. 33;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           || |||||||| | || ||||||
Db    4124 ATCCCCATCAGGATTCCTGCAGA 4102


RESULT   9
US-08-184-012C-5/c
; Sequence 5, Application US/08184012C
; Patent No. 5606029
;  GENERAL INFORMATION:
;    APPLICANT:  Degen, Sandra J. F.
;    TITLE OF INVENTION:  Gene for a growth factor and its cDNA and
```

```
;    TITLE OF INVENTION:  protein
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gregory Lunn
;      STREET:  Wood, Herron & Evans, 2700 Carew Tower
;      CITY:  Cincinnati
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  45202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,  800 Kb
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.5.2
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,012C
;      FILING DATE:  1/18/94
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Gregory
;      REGISTRATION NUMBER:  29,945
;      REFERENCE/DOCKET NUMBER:  CMC 57
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 241-2324
;      TELEFAX:  (513) 421-7269
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6751 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ANTI-SENSE:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;      STRAIN:  Balb/c
;      DEVELOPMENTAL STAGE:  adult
;      TISSUE TYPE:  liver
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic
;      CLONE:  MGL5-12
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  mouse 9, Hgfl locus
;      MAP POSITION:  Trf-Gnai-2-Hgfl-Cck
;    FEATURE:
;      IDENTIFICATION METHOD:  experimental
;    PUBLICATION INFORMATION:
;      RELEVANT RESIDUES IN SEQ ID NO:  5: 1 TO 6751
US-08-184-012C-5


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 6751;
  Best Local Similarity   82.6%;  Pred. No. 33;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           || |||||||| | || ||||||
Db    4124 ATCCCCATCAGGATTCCTGCAGA 4102


RESULT  10
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-20


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 246240;
  Best Local Similarity   82.6%;  Pred. No. 63;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           |||||||| || || ||||||
Db  19868 ATTCCCATGAGGAAACTTGCATA 19890


RESULT  11
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-21


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 246240;
  Best Local Similarity   82.6%;  Pred. No. 63;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           |||||||| || || |||||| |
Db   19868 ATTCCCATGAGGAAACTTGCATA 19890


RESULT  12
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-22


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 246240;
  Best Local Similarity   82.6%;  Pred. No. 63;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           |||||||| || || |||||| |
Db   19868 ATTCCCATGAGGAAACTTGCATA 19890


RESULT  13
US-08-547-182-1/c
; Sequence 1, Application US/08547182
; Patent No. 5672694
;  GENERAL INFORMATION:
;    APPLICANT:  Campbell, Kevin P.
;    APPLICANT:  Lim, Leland
;    APPLICANT:  Duclos, Franck
;    APPLICANT:  Sunada, Yoshihide
;    APPLICANT:  Beckmann, Jacques S.
;    APPLICANT:  Broux, Odile
;    APPLICANT:  Tome, Fernando M.S.
;    APPLICANT:  Fardeau, Michel
;    APPLICANT:  Jackson, Charles E.
;    TITLE OF INVENTION:  -SARCOGLYCAN NUCLEIC ACID SEQUENCE,
;    TITLE OF INVENTION:  DELETERIOUS MUTATIONS AND APPLICATIONS
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/547,182
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  UIRF-9502
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  207-363-0558
;      TELEFAX:  207-363-0528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  10..966
US-08-547-182-1


  Query Match             64.8%;  Score 16.2;  DB 1;  Length 1225;
  Best Local Similarity   85.7%;  Pred. No. 36;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 TCCCATCAGCAATCTTGCAGA 25
           |||||||||||  | ||||||
Db     889 TCCCATCAGCACACATGCAGA 869


RESULT  14
US-08-427-640-28/c
; Sequence 28, Application US/08427640
; Patent No. 5658788
;  GENERAL INFORMATION:
;    APPLICANT:  Berg et al.
;    TITLE OF INVENTION:  Compounds and Methods for Treatment of Thromboembolic Disor
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  IN.
;      COUNTRY:  U.S.A.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.0 Mb storage
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,640
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/689,410
;      FILING DATE:  22 APRIL 1991
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-427-640-28


  Query Match             64.0%;  Score 16;  DB 1;  Length 46;
  Best Local Similarity   79.2%;  Pred. No. 24;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           | || |||||||| | || |||||
Db      39 GCATCCCCATCAGGATTCCTGCAG 16


RESULT  15
US-08-713-569-4/c
; Sequence 4, Application US/08713569
; Patent No. 6040185
;  GENERAL INFORMATION:
;    APPLICANT:  Dietrich, Paul S
;    APPLICANT:  Van Grinsven, Martinus QJM
;    APPLICANT:  Gielen, Johannes JL
;    APPLICANT:  De Haas, Johannes M
;    APPLICANT:  Van Driel, Roeland
;    TITLE OF INVENTION:  Genetic Stabilizing Elemnents
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SANDOZ AGRO INC.
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/713,569
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/439,156
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marcus-Wyner, Lynn
;      REGISTRATION NUMBER:  34,869
;      REFERENCE/DOCKET NUMBER:  135-1086
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)354-3588
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  862 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Zea mays
US-08-713-569-4


  Query Match             64.0%;  Score 16;  DB 5;  Length 862;
  Best Local Similarity   79.2%;  Pred. No. 42;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAGA 25
           |||||||||||||| | ||   |||
Db      85 AATTCCCATCAGCTACCTACGAGA 62


Search completed: June 22, 2000, 15:21:54
Job time: 18792 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:43 ; Search time 458.59 Seconds
(without alignments)
13.639 Million cell updates/sec

Title: US-09-362-485-14
Perfect score: 25
Sequence: 1 GAATTCCCATCAGCAATCTTGCAGA 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters. 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	25	1	V49515	Mycobacterium sp.
2		25	100.0	25	1	V49613	AlaDH derived olig
3		25	100.0	1260	1	V49510	Mycobacterium sp.
4		25	100.0	1260	1	V49625	Mycobacterium tube
c 5		17	68.0	253	1	V89998	EST clone CW835. N
6		17	68.0	2465	1	Q13136	Hyp operon. DNA en
c 7		17	68.0	4570	1	T01190	Barbary duck parvo
c 8		17	68.0	11307	1	X12986	Enterococcus faeca
9		16.6	66.4	607	1	X40367	Human secreted pro
10		16.6	66.4	807	1	T10196	Human elastase IV
11		16.6	66.4	807	1	T15536	DNA encoding human
12		16.6	66.4	959	1	V11288	Human elastase hom
13		16.6	66.4	959	1	V82012	Human homologue of
c 14		16.6	66.4	2188	1	Q79726	Mouse L5/3 tumour
c 15		16.6	66.4	2188	1	T62441	Mouse growth facto
c 16		16.6	66.4	2216	1	V72084	Mouse MSP cDNA. Tr
c 17		16.6	66.4	2281	1	V57909	Human haemochromat
c 18		16.6	66.4	6751	1	Q79727	Mouse L5/3 tumour
c 19		16.6	66.4	6751	1	T62442	Mouse growth facto
20		16.6	66.4	7147	1	V52156	Streptococcus pneu
c 21		16.6	66.4	235033	1	V57926	Hereditary haemoch
c 22		16.6	66.4	237326	1	V57903	Hereditary haemoch
c 23		16.4	65.6	1236	1	X20633	Polynucleotide seq
c 24		16.2	64.8	1225	1	T89303	Human beta-sarcogl
c 25		16.2	64.8	1225	1	T89304	Human beta-sarcogl
c 26		16.2	64.8	1225	1	T89291	Human beta-sarcogl
c 27		16.2	64.8	1225	1	T89305	Human beta-sarcogl
28		16.2	64.8	1380	1	Q84044	Japonicum allergen
29		16.2	64.8	1395	1	Q66050	Japanese cedar pol
30		16.2	64.8	1410	1	Q66049	Japanese cedar pol
31		16.2	64.8	1479	1	Q66051	Japanese cedar pol
32		16.2	64.8	1542	1	Q84045	Japonicum allergen
33		16.2	64.8	1726	1	Q66048	Japanese cedar pol
34		16.2	64.8	1733	1	Q84046	Japonicum allergen
35		16.2	64.8	1733	1	T18102	Japan cedar pollen
c 36		16.2	64.8	2037	1	V15736	Beta-sarcoglycane
37		16	64.0	359	1	V75340	Staphylococcus aur
c 38		16	64.0	862	1	Q78182	Maize associated r
c 39		16	64.0	1068	1	T62607	DNA encoding t-PA
c 40		16	64.0	1068	1	T87457	mt-PA9. Recombinan
c 41		16	64.0	2200	1	T16309	Arabidopsis CRY2 g
c 42		16	64.0	3188	1	Q66685	Delta(4)-3-ketoste
c 43		16	64.0	28720	1	V49655	Human SC3 DNA. Pro
c 44		15.8	63.2	2003	1	T59619	G-protein parathyr
c 45		15.8	63.2	2182	1	V00731	Pea plastidial pho

ALIGNMENTS

RESULT 1
V49515
ID V49515 standard; DNA; 25 BP.
AC V49515;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F4.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 25 BP; 8 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 25; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCCCATCAGCAATCTTGCAGA 25
||||||||||||||||||||||||||
Db 1 GAATTCCCATCAGCAATCTTGCAGA 25

RESULT 2
V49613
ID V49613 standard; DNA; 25 BP.
AC V49613;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F4.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 25 BP; 8 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 25; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCCCATCAGCAATCTTGCAGA 25
||||||||||||||||||||||||||
Db 1 GAATTCCCATCAGCAATCTTGCAGA 25

RESULT 3
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 25; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCCCATCAGCAATCTTGCAGA 25
||||||||||||||||||||||||||
Db 1 GAATTCCCATCAGCAATCTTGCAGA 25

RESULT 4
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 25; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCCCATCAGCAATCTTGCAGA 25
||||||||||||||||||||||||||
Db 1 GAATTCCCATCAGCAATCTTGCAGA 25

RESULT 5
V89998/c
ID V89998 standard; cDNA; 253 BP.
AC V89998;
DT 15-FEB-1999 (first entry)
DE EST clone CW835.
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
OS Homo sapiens.
PN WO9845436-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06955.
PR 10-APR-1997; US-838821.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070077/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 398; 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make

```
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
SQ   Sequence   253 BP;    75 A;    60 C;    48 G;    70 T;

  Query Match             68.0%;  Score 17;  DB 1;  Length 253;
  Best Local Similarity   80.0%;  Pred. No. 34;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || |||||||||  | |||||||  |
Db     165 GATTTCCCATCAAAACTCTTGCAAA 141


RESULT   6
Q13136
ID   Q13136 standard; DNA; 2465 BP.
AC   Q13136;
DT   22-OCT-1991  (first entry)
DE   Hyp operon.
KW   Antibodies; heat shock; hypersensitive; allergen; HSP60; GroEL;
KW   GroES; ss.
OS   Chlamydia psittaci GPIC.
FH   Key             Location/Qualifiers
FT   cds             314. .622
FT                   /*tag= a
FT                   /label= hypA
FT   cds             673. .2307
FT                   /*tag= b
FT                   /label= hypB
FT   -35_signal      83. .88
FT                   /*tag= c
FT   -10_signal      120. .127
FT                   /*tag= d
FT   rbs             301. .304
FT                   /*tag= e
FT   rbs             656. .659
FT                   /*tag= f
FT   terminator      2404. .2440
FT                   /*tag= g
FT                   /note=" dyad symmetry; rho-independent"
FT   misc_feature    350. .385
FT                   /*tag= h
FT                   /label= oligo
FT                   /note= "used for Northern blot analysis"
FT   misc_feature    2152. .2187
FT                   /*tag= i
FT                   /label= oligo
FT                   /note= "used in Northern blot analysis"
PN   US7531317-A.
PD   09-JUL-1991.
PF   31-MAY-1990; 143560.
PR   31-MAY-1990; US-531317.
PA   (USSH ) NAT INST OF HEALTH.
DR   WPI; 91-245693/33.
DR   P-PSDB; R13334.
DR   P-PSDB; R13335.
PT   DNA encoding HypA and HypB Chlamydia proteins - used to develop
PT   prods. for detection of and vaccines against Chlamydia infection.
PS   Disclosure; Fig 5; 51pp; English.
CC   The sequence was obtd. from clone pGP57 isolated from a DNA library
CC   prepd. from C. psittaci genomic DNA.   The two ORFs encode HypA and
CC   HypB hypersensitivity proteins, regulated by a heat shock type
CC   promoter region. The DNA can be used to produce recombinant
CC   proteins or to design probes for the detection of Chlamydial
CC   infection. The Hyp proteins can be used to to raise antibodies and
CC   in vaccines. The Hyp B protein elicits a cell-mediated immune
CC   response so can be used as a skin test antigen.
CC   See also Q13137.
SQ   Sequence   2465 BP;    853 A;    467 C;    509 G;    636 T;

  Query Match             68.0%;  Score 17;  DB 1;  Length 2465;
  Best Local Similarity   80.0%;  Pred. No. 46;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || |||||||  || ||||||  |||
Db      66 GACTTCCCATTAGAAATCTTCAAGA 90


RESULT   7
T01190/c
ID   T01190 standard; DNA; 4570 BP.
AC   T01190;
DT   09-MAY-1996  (first entry)
DE   Barbary duck parvovirus fragment.
KW   Barbary duck; parvovirus; open reading frame; non-structural; VP1; VP2;
KW   liver extract; virus particle; vector; host cell; attenuated; geese; ss.
OS   Barbary duck parvovirus.
FH   Key             Location/Qualifiers
FT   cds             318. .2198
FT                   /*tag= a
FT                   /product= parvovirus non-structural protein
FT                   /note= "open reading frame 1"
FT   cds             954. .956
FT                   /*tag= b
FT                   /transl_except= Seq: GTA, a.a.: Ala
FT   cds             2217. .4415
FT                   /*tag= c
FT                   /produc = parvovirus VP1 protei
FT                   /note=  pen reading frame 2"
FT   cds             2811. .4415
FT                   /*tag= d
FT                   /product= parvovirus VP2 protein
FT                   /note= "open reading frame 3"
PN   FR2718151-A1.
PD   06-OCT-1995.
PF   03-MAR-1995; 002515.
PR   03-MAR-1994; HU-000634.
PA   (SNFI ) PHYLAXIA-SANOFI OLTOANYAGTERMELO RT.
PI   Erdei J,  Kisari J,  Nagy J,  Zadori Z;
DR   WPI; 95-346564/45.
DR   P-PSDB; R85384, R85385, R85386.
PT   Nucleic acid sequence from Barbary duck parvovirus - related
PT   vectors, transformed cells, proteins etc., useful diagnostically and
PT   in protective vaccines
PS   Claim 1; Fig 2; 39pp; French.
CC   The sequence of a fragment from the Barbary duck parvovirus strain FM,
CC   which contains 3 open reading frames (ORF), with ORF3 being internal to
CC   ORF2, encoding a non-structural, the VP1 and VP2 proteins respectively.
CC   The virus was isolated from a 13 day old duck that had died of Barbary
CC   duck parvovirus infection, by incubating liver extracts injected into
CC   embryonated duck eggs.  The DNA was isolated from the resulting 20-25 nm
CC   viral particles.  The DNA sequences encoding the non-structural (R85384),
CC   VP1 (R85385) or VP2 (R85386) proteins can be used to construct viral
CC   vectors, host cells or attenuated parvoviruses for use as vaccines to
CC   protect Barbary ducks and geese.
SQ   Sequence   4570 BP;    1424 A;    983 C;    1057 G;    1106 T;

  Query Match             68.0%;  Score 17;  DB 1;  Length 4570;
  Best Local Similarity   80.0%;  Pred. No. 50;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           ||||| |||||  ||||| |||| |
Db    1615 GAATTTCCATCCACAATCATGCACA 1591
```

```
RESULT   8
X12986/c
ID   X12986 standard; DNA; 11307 BP.
AC   X12986;
DT   19-MAR-1999  (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO:49.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; U08985.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 444-450; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence   11307 BP;    3719 A;    1967 C;    2462 G;    3144 T;

  Query Match             68.0%;  Score 17;  DB 1;  Length 11307;
  Best Local Similarity   80.0%;  Pred. No. 56;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           | | | ||||||||||||| ||| ||
Db    5378 GTAATTCCATCAGCAATCATGCCGA 5354


RESULT   9
X40367
ID   X40367 standard; cDNA; 607 BP.
AC   X40367;
DT   16-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID NO:154.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906439-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1233.
PR   01-AUG-1997; US-904468.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153700/13.
DR   P-PSDB; Y11649.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from liver, lung, large intestine, colon,
PT   thyroid and pancreas tissue
PS   Claim 1; Page 264-266; 398pp; English.
CC   X40251 to X40397 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y11533 to Y11679,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   607 BP;    111 A;    184 C;    178 G;    116 T;

  Query Match             66.4%;  Score 16.6;  DB 1;  Length 607;
  Best Local Similarity   82.6%;  Pred. No. 58;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| |||||||||
Db     386 ATTGCCCTCATCAAGCTTGCAGA 408


RESULT  10
T10196
ID   T10196 standard; cDNA; 807 BP.
AC   T10196;
DT   05-SEP-1996  (first entry)
DE   Human elastase IV gene.
KW   Human; elastase IV; pancreas; amplification; primer; PCR; pQE-9; E.coli;
KW   polymerase chain reaction; expression vector; nickel chelate; elasticity;
KW   HexaHis tag; arterial wall; serum lipid; lipoprotein; metabolism;
KW   antagonist; septic shock; inflammation; arthritis; pancreatitis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  1. .48
FT                   /*tag= a
FT   misc_feature    49. .87
FT                   /*tag= b
FT                   /note= "encodes prosequence"
FT   mat_peptide     88. .804
FT                   /*tag= c
FT   misc_difference 475. .477
FT                   /*tag= d
FT                   /transl_except= seq: GGC, a.a.: Glu
FT   misc_difference 481. .483
FT                   /*tag= e
FT                   /transl_except= seq: GGC, a.a.: Glu
FT   misc_difference 499. .501
FT                   /*tag= f
FT                   /transl_except= seq: GGC, a.a.: Glu
PN   WO9601270-A1.
PD   18-JAN-1996.
PF   06-JUN-1995; U07236.
PR   05-JUL-1994; US-270584.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Adams MD,  Greene JM;
DR   WPI; 96-087629/09.
DR   P-PSDB; R88481.
PT   New nucleic acid encoding human elastase IV - used to compensate
PT   elastase deficiency and to identify specific antagonists for
PT   treating inflammation etc.
PS   Claim 1; Page 36-37; 46pp; English.
CC   This is the nucleotide sequence of the human elastase IV gene. The
CC   gene was isolated from a cDNA library derived from human pancreatic
```

```
CC   cells.  The sequence encoding the mature protein (amino acids 14-252)
CC   was amplified using the primers T27099-T27100.  The amplified region
CC   was cloned into plasmid pQE-9 for expression in E.coli and purification
CC   on a Nickel chelate column using the HexaHis tag.  The protein is
CC   used to treat conditions requiring elastase IV e.g. to restore
CC   elasticity of arterial walls, improve serum lipid levels and serum
CC   lipoprotein metabolism.  Antagonists of elastase IV can be used to
CC   treat e.g. septic shock, inflammation, arthritis or pancreatitis.
SQ   Sequence   807 BP;    159 A;    243 C;    248 G;    157 T;


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 807;
  Best Local Similarity   82.6%;  Pred. No. 60;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     364 ATTGCCCTCATCAAGCTTGCAGA 386


RESULT  11
T15536
ID   T15536 standard; cDNA to mRNA; 807 BP.
AC   T15536;
DT   24-JUL-1997  (revised)
DT   11-NOV-1996  (first entry)
DE   DNA encoding human caldecrin.
KW   caldecrin; rat; human; calcium serum lowering agent;
KW   recombinant production; preprosequence; ds.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   misc_feature    1. .39
FT                   /*tag=  a
FT                   /note= "pre-pro-sequence"
FT   misc_feature    46. .87
FT                   /*tag=  b
FT                   /note= "pro-sequence"
FT   mat_peptide     88. .804
FT                   /*tag=  c
FT                   /product= mature_caldecrin
FT   misc_difference 238
FT                   /*tag=  d
FT                   /note= "T or C"
PN   WO9600287-A.
PD   04-JAN-1996.
PF   23-JUN-1995; J01268.
PR   24-JUN-1994; JP-164898.
PR   08-MAR-1995; JP-074676.
PA   (CHUS ) CHUGAI SEIYAKU KK.
PI   Akiyama M,  Noikura T,  Saheki T,  Tomomura A;
DR   WPI; 96-068871/07.
DR   P-PSDB; R90682.
PT   DNA encoding rat and human caldecrin and related vectors - for prodn.
PT   of caldecrin for use as serum calcium lowering agent
PS   Claim 11; Page 26-27; 43pp; Japanese.
CC   The DNA encodes human caldecrin preprosequence (see R90683). Expression
CC   vectors contg. the DNA can be used for the recombinant production of
CC   caldecrin from transformant host cells. The caldecrin is used as a
CC   serum calcium lowering agent.
CC   (Revised entry submitted to correct sequence analysis breakdown.)
SQ   Sequence   807 BP;    159 A;    243 C;    247 G;    157 T;


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 807;
  Best Local Similarity   82.6%;  Pred. No. 60;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     364 ATTGCCCTCATCAAGCTTGCAGA 386


RESULT  12
V11288
ID   V11288 standard; cDNA; 959 BP.
AC   V11288;
DT   15-JUL-1998  (first entry)
DE   Human elastase homologue HEIV cDNA.
KW   Elastase; HEIV; human; pancreas; serine protease; tissue wasting;
KW   diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             22. .828
FT                   /*tag=  a
FT                   /product= heIV
FT                   /note= "elastase"
PN   US5738991-A.
PD   14-APR-1998.
PF   06-DEC-1995; 568031.
PR   06-DEC-1995; US-568031.
PA   (INCY-) INCYTE PHARM INC.
PI   Braxton SM, Delegeane AM, Diep D;
DR   WPI; 98-250414/22.
DR   P-PSDB; W40530.
PT   DNA encoding human homologue of rat elastase IV - is useful for the
PT   recombinant production of the protein and hybridisation probes for
PT   detecting sequences encoding the protein
PS   Claim 2; Fig 1; 19pp; English.
CC   This cDNA sequence encodes a novel human homologue of rat elastase
CC   IV (HEIV) protein derived from the pancreas. Excessive serine protease
CC   production (including elastase) results in tissue wasting. Antisense
CC   molecules to this protein are useful for hybridising to DNA in samples
CC   to diagnose abnormalities in gene expression associated with tissue
CC   wasting. HEIV and its related polynucleotides and proteins are novel
CC   products for various uses with elastase related conditions.
SQ   Sequence   959 BP;    200 A;    270 C;    296 G;    193 T;


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 959;
  Best Local Similarity   82.6%;  Pred. No. 61;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||| || ||| ||| ||||||||
Db     385 ATTGCCCTCATCAAGCTTGCAGA 407


RESULT  13
V82012
ID   V82012 standard; cDNA; 959 BP.
AC   V82012;
DT   09-MAR-1999  (first entry)
DE   Human homologue of rat elastase IV encoding cDNA.
KW   Human; elastase IV; HEIV; tissue wasting; oedema; vascular damage;
KW   interstitial haemorrhage; coagulation necrosis; fat necrosis;
KW   parenchymal cell necrosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             22. .828
FT                   /*tag=  a
PN   US5856109-A.
PD   05-JAN-1999.
PF   07-NOV-1997; 966319.
PR   06-DEC-1995; US-568031.
PR   07-NOV-1997; US-966319.
PA   (INCY-) INCYTE PHARM INC.
PI   Braxton SM, Delegeane AM, Diep D;
DR   WPI; 99-105100/09.
DR   P-PSDB; W89410.
PT   Method for producing polypeptide New isolated human homologue of rat
PT   elastase IV - used to develop products for the diagnosis, prevention
PT   and treatment of disease states or conditions associated with
PT   excessive tissue wasting
PS   Example; Column 23-24; 19pp; English.
CC   A method has been developed for producing a polypeptide comprising: (a)
CC   culturing a host cell transformed with an expression vector comprising a
```

```
CC   polynucleotide (PN) encoding a human homologue of rat elastase IV (HEIV)
CC   for the expression of the protein; and (b) recovering the protein from a
CC   host cell culture. The present sequence encodes HEIV. Products from the
CC   present invention can be used to identify inhibitors of HEIV activity.
CC   HEIV inhibitors can be used to treat or prevent excessive tissue wasting,
CC   e.g. oedema, interstitial haemorrhage, vascular damage, coagulation
CC   necrosis, fat necrosis, and parenchymal cell necrosis. The products can
CC   also be used for detection, diagnosis and drug screening.
SQ   Sequence   959 BP;    200 A;    270 C;    296 G;    193 T;


  Query Match            66.4%;  Score 16.6;  DB 1;  Length 959;
  Best Local Similarity  82.6%;  Pred. No. 61;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 ATTCCCATCAGCAATCTTGCAGA 25
          ||| || ||| ||| ||||||||
Db    385 ATTGCCCTCATCAAGCTTGCAGA 407


RESULT  14
Q79726/c
ID   Q79726 standard; cDNA; 2188 BP.
AC   Q79726;
DT   13-FEB-1995  (first entry)
DE   Mouse L5/3 tumour suppressor gene.
KW   Mouse L5/3 gene; small lung cell carcinoma; tumour suppression;
KW   chromosome 3; 3p21; D3F15S2 locus; hepatocyte growth factor;
KW   renal cell carcinoma; Von Hipple-Lindau syndrome; predisposition;
KW   kringle domain; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             1. .2107
FT                   /*tag= a
FT                   /note= "the first nucleotide of this sequence
FT                   represents the last position of codon 19"
FT   variation       12
FT                   /*tag= b
FT                   /note= "CCG (Pro) codon in Q79726 (cDNA); CAG (Gln)
FT                   codon in Q79727 (genomic DNA). This site is
FT                   anticipated to be polymorphic in the
FT                   population so that both versions encode
FT                   functional L5/3 protein"
PN   US5315000-A.
PD   24-MAY-1994.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
PA   (CHIL-) CHILDRENS HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 94-166645/20.
DR   P-PSDB; R66600.
PT   DNA from D3F15S2 locus of human chromosome 3 - encoding novel
PT   growth factor, L5/3, useful as probe for detecting
PT   pre-deposition towards cancer
PS   Disclosure; Columns 27-32; 31pp; English.
CC   A mouse liver cDNA library was screened with cDNA coding for human
CC   L5/3; the L5/3 gene, located at the D3F15S2 locus of human
CC   chromosome 3, codes for a protein composed of 4 kringle domains,
CC   followed by a serine protease-like domain. The longest murine clone
CC   to be isolated (pML5-2, Q79726) was not full-length. The open
CC   reading frame was present at the 5'end of the sequence with no
CC   codon for the initiator methionine in-frame with the coding
CC   sequence. After determination of the sequence of the mouse gene it
CC   was determined that the cDNA lacked 44bp of coding and 94bp of
CC   non-coding sequence at its 5'-end.
SQ   Sequence   2188 BP;    509 A;    608 C;    627 G;    444 T;


  Query Match            66.4%;  Score 16.6;  DB 1;  Length 2188;
  Best Local Similarity  82.6%;  Pred. No. 68;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 ATTCCCATCAGCAATCTTGCAGA 25
          || |||||||| | || ||||||
Db   1258 ATCCCCATCAGGATTCCTGCAGA 1236


RESULT  15
T62441/c
ID   T62441 standard; cDNA to mRNA; 2188 BP.
AC   T62441;
DT   21-JUL-1997  (first entry)
DE   Mouse growth factor L5/3 partial cDNA clone ML5-2.
KW   Mouse; growth factor; foetal; liver; probe; bovine; prothrombin; locus;
KW   polymorphism; transition; exon; intron; chromosome; kringle domain; ss;
KW   cell growth; tumour suppressor; hepatocyte growth factor; regeneration.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             2. .2107
FT                   /*tag= a
FT                   /product= partial mouse growth factor L5/3
FT   signal_peptide  2. .49
FT                   /*tag= b
FT                   /note= "partial coding sequence for signal peptide"
FT   mat_peptide     50. .2104
FT                   /*tag= c
PN   US5606029-A.
PD   25-FEB-1997.
PF   14-MAY-1992; 882925.
PR   14-MAY-1992; US-882925.
PR   18-JAN-1994; US-184012.
PA   (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
PI   Degen SJ;
DR   WPI; 97-153621/14.
DR   P-PSDB; W14271.
PT   Human growth factor protein L5/3 - useful for altering cell growth,
PT   e.g. as tumour suppressor
PS   Disclosure; Column 27-32; 34pp; English.
CC   This is the nucleotide sequence of the insert isolated from clone ML5-2
CC   which encodes part of a mouse growth factor designated L5/3 (W14272).
CC   This sequence was isolated from a lambda-gt10 mouse liver cDNA library
CC   using a fragment of the corresponding human cDNA sequence (T62436) as a
CC   probe.  The fragment presented here is a partial sequence which lacks
CC   44 bp of the coding region and 94 bp of the non-coding region.  A 1450 bp
CC   fragment of this sequence (nucleotides 738-2188) covering 8 amino acids
CC   of the second kringle domain and all of the third and fourth kringle
CC   domains and the serine protease-like domain, was used to screen a mouse
CC   liver genomic DNA library to isolate the full length genomic sequence
CC   (T62442).  The protein can be used to alter cell growth (as a growth
CC   factor or tumour suppressor) and has similar properties to the hepatocyte
CC   growth factor that is actively involved in liver regeneration.
SQ   Sequence   2188 BP;    509 A;    608 C;    627 G;    444 T;


  Query Match            66.4%;  Score 16.6;  DB 1;  Length 2188;
  Best Local Similarity  82.6%;  Pred. No. 68;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 ATTCCCATCAGCAATCTTGCAGA 25
          || |||||||| | || ||||||
Db   1258 ATCCCCATCAGGATTCCTGCAGA 1236


Search completed: June 22, 2000, 15:12:46
Job time: 18092 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:00 ; Search time 8627.09 Seconds
(without alignments)
-2.819 Million cell updates/sec

Title: US-09-362-485-14
Perfect score: 25
Sequence: 1 GAATTCCCATCAGCAATCTTGCAGA 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	25	5	A87617	A87617 Sequence 14
2		25	100.0	1260	5	A87604	A87604 Sequence 1
3		25	100.0	1260	5	A89744	A89744 Sequence 1
4		25	100.0	2584	1	MTALADH	X63069 M.tuberculo
5		21.8	87.2	56414	1	MTV002	AL008967 Mycobacte
6		19.8	79.2	180588	45	AC010703	AC010703 Homo sapi
7		19.2	76.8	165737	44	AC018616	AC018616 Homo sapi
8	c	19.2	76.8	172513	4	AC011958	AC011958 Homo sapi
9		19.2	76.8	176147	45	AC020694	AC020694 Homo sapi
10		19.2	76.8	189642	32	AL135797	AL135797 Homo sapi
11	c	19.2	76.8	198436	33	AL133553	AL133553 Homo sapi
12	c	18.6	74.4	1401	3	OAKRT211	X73455 Ovis aries
13		18.6	74.4	2468	1	CPHYPAB	X51404 Chlamydophi
14	c	18.6	74.4	141602	52	AC013732	AC013732 Homo sapi
15	c	18.6	74	172175	33	HSJ831A12	AL122019 Homo sapi
16	c	18.4	73	177756	33	AC005020	AC005020 Homo sapi
17	c	18.4	73.6	197273	51	AC002461	AC002461 Human BAC
18	c	18.2	72.8	417	11	HUMDMDJ	L37425 Homo sapien
19		18.2	72.8	1665	7	YSCSDH	J05487 S.cerevisia
20	c	18.2	72.8	10130	7	SCYLL040C	Z73145 S.cerevisia
21		18.2	72.8	12215	8	AF001317	AF001317 Saccharom
22	c	18.2	72.8	16966	34	CEY55D9A	AL032649 Caenorhab
23		18.2	72.8	54708	50	AC007209	AC007209 Arabidops
24	c	18.2	72.8	115355	32	CEY55D9	AL008876 Caenorhab
25		18.2	72.8	123400	33	AL135934	AL135934 Homo sapi
26	c	18.2	72.8	146388	40	AC005880	AC005880 citb_43_a
27	c	18.2	72.8	148608	32	AP001124	AP001124 Homo sapi
28	c	18.2	72.8	151490	43	AC011883	AC011883 Homo sapi
29		18.2	72.8	154796	40	AC007372	AC007372 Homo sapi
30	c	18.2	72.8	161598	43	AC016801	AC016801 Homo sapi
31	c	18.2	72.8	168677	11	CNS0000F	AL049837 Human chr
32		18.2	72.8	173360	44	AC013362	AC013362 Homo sapi
33		18.2	72.8	176929	40	AC004887	AC004887 Homo sapi
34		18.2	72.8	178317	41	AC010929	AC010929 Homo sapi
35	c	18.2	72.8	186593	32	AP001001	AP001001 Homo sapi
36		18.2	72.8	186666	43	AC012564	AC012564 Homo sapi
37	c	18.2	72.8	215640	1	BSUB0016	Z99119 Bacillus su
38		18.2	72.8	220060	2	AF008220	AF008220 Bacillus
39	c	18.2	72.8	247661	57	AC012154	AC012154 Homo sapi
40		17.8	71.2	86574	11	HS833B7	AL008637 Human DNA
41	c	17.8	71.2	104147	11	AC004703	AC004703 Homo sapi
42	c	17.8	71.2	147478	33	AL139334	AL139334 Homo sapi
43	c	17.8	71.2	150052	44	AC016535	AC016535 Homo sapi
44		17.8	71.2	181822	33	AL139256	AL139256 Homo sapi
45	c	17.8	71.2	200946	53	AC018501	AC018501 Homo sapi

ALIGNMENTS

```
RESULT   1
A87617
LOCUS       A87617         25 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 14 from Patent WO9836089.
ACCESSION   A87617
VERSION     A87617.1  GI:6736257
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 25)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .25
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        8 a      7 c      4 g      6 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 5;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 0.025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGA 25
          |||||||||||||||||||||||||
Db      1 GAATTCCCATCAGCAATCTTGCAGA 25


RESULT   2
A87604
LOCUS       A87604       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9836089.
ACCESSION   A87604
VERSION     A87604.1  GI:6736244
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGA 25
          |||||||||||||||||||||||||
Db      1 GAATTCCCATCAGCAATCTTGCAGA 25


RESULT   3
A89744
LOCUS       A89744       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9832862.
ACCESSION   A89744
VERSION     A89744.1  GI:6738278
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGA 25
          |||||||||||||||||||||||||
Db      1 GAATTCCCATCAGCAATCTTGCAGA 25


RESULT   4
MTALADH
LOCUS       MTALADH      2584 bp    DNA             BCT       01-DEC-1993
DEFINITION  M.tuberculosis gene for L-alanine dehydrogenase.
ACCESSION   X63069 S36765
VERSION     X63069.1  GI:44565
KEYWORDS    alanine dehydrogenase; extracellular; intracellular.
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 2584)
  AUTHORS   Andersen,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1991) A.B. Andersen, Statens Seruminstitut,
            Artillerivej 5, DK-2300 Copenhagen S, DENMARK
REFERENCE   2  (bases 1 to 2584)
  AUTHORS   Andersen,A.B., Andersen,P. and Ljungqvist,L.
  TITLE     Structure and function of a 40,000-molecular-weight protein antigen
            of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 60 (6), 2317-2323 (1992)
  MEDLINE   92267644
FEATURES             Location/Qualifiers
     source          1. .2584
                     /organism="Mycobacterium tuberculosis"
                     /strain="Erdman"
                     /isolate="TMC 107, lot 9A-2"
                     /db_xref="taxon:1773"
                     /clone_lib="lambda gt11"
                     /clone="lambda AA67"
     CDS             76. .1197
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="alanine  dehydrogenase"
                     /protein_id="CAA44791.1"
                     /db_xref="GI:44566"
                     /db_xref="SWISS-PROT:P30234"
                     /translation="MRVGIPTETKNNEFQFRVAITPAGVAELTRRGHEVLIQAGAGEG
                     SAITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAAS
                     RACTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGV
                     LMGGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIH
                     TRYSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGC
                     FEGSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAA
                     CRSNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      463 a    871 c    819 g    431 t
ORIGIN
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 2584;
  Best Local Similarity   100.0%;  Pred. No. 0.031;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           |||||||||||||||||||||||||
Db       1 GAATTCCCATCAGCAATCTTGCAGA 25


RESULT   5
MTV002
LOCUS       MTV002      56414 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
ACCESSION   AL008967 AL123456
VERSION     AL008967.1  GI:3261491
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 56414)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 56414)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2624256.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .56414
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .>56414
                     /note="fragment designated v002. Does not represent a
                     physical clone"
     source          <1. .233
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y154"
     gene            complement(3. .527)
                     /gene="recX"
     CDS             complement(3. .527)
                     /gene="recX"
                     /note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar
                     to eg. RECX_PSEAE P37860 regulatory protein recx from
                     Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161
                     z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
                     overlap. Overlaps and extends CDS from ovelapping cosmid
                     MTCY154.16c"
                     /codon_start=1
                     /transl_table=11
                     /product="recX"
                     /protein_id="CAA15532.1"
                     /db_xref="GI:2624257"
                     /db_xref="SPTREMBL:O33280"
                     /translation="MTVSCPPPSTSEREEQARALCLRLLTARSRTRAELAGQLAKRGY
                     PEDIGNRVLDRLAAVGLVDDTDFAEQWVQSRRANAAKSKRALAAELHAKGVDDDVITT
                     VLGGIDAGAERGRAEKLVRARLRREVLIDDGTDEARVSRRLVAMLARRGYGQTLACEV
                     VIAELAAERERRRV"
     gene            complement(493. .2865)
                     /gene="recA"
     CDS             complement(493. .2865)
                     /gene="recA"
                     /note="Rv2737c, (MTV002.02c), len: 790 aa. recA,
                     identicalto RECA_MYCTU P26345 reca protein (790 aa).
                     Contains self-splicing protein element (intein) from 2294
                     to 969 (c), similar to intein II from TR:E332317
                     (EMBL:Y13030) DNA-directedDNA polymerase (EC 2.7.7.7) from
                     Thermococcus sp. (1829 aa), fasta scores; opt: 81 z-score:
                     235.2 E(): 6e-06, 24.6%identity in 183 aa overlap.
                     Contains PS00017 ATP/GTP-bindingsite motif A (P-loop),
                     PS00321 recA signature, and PS00881Protein splicing
                     signature. See Davis et al, (1992) Cell71(2):201-210"
                     /codon_start=1
                     /transl_table=11
                     /product="recA"
                     /protein_id="CAA15533.1"
                     /db_xref="GI:2624258"
                     /db_xref="SWISS-PROT:P26345"
                     /translation="MTQTPDREKALELAVAQIEKSYGKGSVMRLGDEARQPISVIPTG
                     SIALDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGVAAFIDAEHALDP
                     DYAKKLGVDTDSLLVSQPDTGEQALEIADMLIRSGALDIVVIDSVAALVPRAELEGEM
                     GDSHVGLQARL SQALRKMTGALNNSGTTAIFINQLRDKIGVMFGSPETTTGGKALKF
                     YASVRMDVRRV TLKDGTNAVGNRTRVKVVKNKCLAEGTRIFDPVTGTTHRIEDVVDG
                     RKPIHVVAAAKDGTLHARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRA
                     AGELRKGDRVAQPRRFDGFGDSAPIPADHARLLGYLIGDGRDGWVGGKTPINFINVQR
                     ALIDDVTRIAATLGCAAHPQGRISLAIAHRPGERNGVADLCQQAGIYGKLAWEKTIPN
                     WFFEPDIAADIVGNLLFGLFESDGWVSREQTGALRVGYTTTSEQLAHQIHWLLLRFGV
                     GSTVRDYDPTQKRPSIVNGRRIQSKRQVFEVRISGMDNVTAFAESVPMWGPRGAALIQ
                     AIPEATQGRRRGSQATYLAAEMTDAVLNYLDERGVTAQEAAAMIGVASGDPRGGMKQV
                     LGASRLRRDRVQALADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTL
                     VAEGVVVHNCSPPFKQAEFDILYGKGISREGSLIDMGVDQGLIRKSGAWFTYEGEQLG
                     QGKENARNFLVENADVADEIEKKIKEKLGIGAVVTDDPSNDGVLPAPVDF"
     misc_feature    complement(790. .2112)
                     /gene="recA"
                     /note="recA intein"
     misc_feature    complement(790. .807)
                     /gene="recA"
                     /note="PS00881 Protein splicing signature"
     misc_feature    complement(2197. .2223)
                     /gene="recA"
                     /note="PS00321 recA signature"
     misc_feature    complement(2644. .2667)
                     /gene="recA"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     RBS             complement(2873. .2877)
                     /note="possible RBS upstream of Rv2737c"
     gene            complement(3247. .3453)
                     /gene="Rv2738c"
```

```
     CDS             complement(3247. .3453)
                     /gene="Rv2738c"
                     /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown,
                     N-terminus is highly similar to the N-terminus of the
                     upstream ORFMTV002.07c (78.4% identity in 37 aa overlap);
                     also similarto AL020958|SC4H8_5 Streptomyces coelicolor
                     cosmid 4H8 (64aa)  opt:  185 z-score: 283.5 E(): 2.9e-08;
                     39.7% identityin 63 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2738c"
                     /protein_id="CAA15534.1"
                     /db_xref="GI:2624260"
                     /db_xref="SPTREMBL:O33281"
                     /translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHVLTGFDGRSAAQAI
                     EDGVEPRDVWRALCADFDVPHDRW"
     gene            complement(3464. .4630)
                     /gene="Rv2739c"
     CDS             complement(3464. .4630)
                     /gene="Rv2739c"
                     /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable
                     transferase, similar to eg. TR:O51560 (EMBL:L28170)
                     Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
                     fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
                     identity in 425 aaoverlap. Equivalent to Mycobacterium
                     leprae protein MLCB33.02c (392 aa); fasta scores
                     gp|Z94723|MLCB33_2 Mycobacteriumleprae cosmid B33 opt:
                     2112 z-score: 2364.5 E(): 0; 80.9%identity in 388 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2739c"
                     /protein_id="CAA15535.1"
                     /db_xref="GI:2624261"
                     /db_xref="SPTREMBL:O33282"
                     /translation="MRVAVVAGPDPGHSFPAIALCQRFRAAADTPTLFTGVEWLEAAR
                     AAGIDAVELDGLAATDRDLDAGARIHRRAAQMAVLNVPRLRALEPELVVSDVITACGG
                     MAAELLGIPWVELNPHPLYLPSKGLPPIGSGLAAGTGIRGRLRDATMRALTGRSWRAG
                     LRQRAAVRVEIGLPARDPGPLRRLIATLPALEVPRPDWPAEAVVVGPLHFEPTDRVLA
                     IPAGTGPVVVVAPSTALTGTAGLTEVALQSLTPGETVPSGSRLVVSRLSGADLTVPPW
                     AVAGLGSQAELLTRADLVICGGGHGMVAKTLLAGVPMVVVPGGGDQWEIANRVVRQGS
                     AVLIRPLTADALVAAVNEVLSSPRFREAARRAAASVAGAADPVRVCHDALALAG"
     gene            4674. .5123
                     /gene="Rv2740"
     CDS             4674. .5123
                     /gene="Rv2740"
                     /note="Rv2740, (MTV002.05), len: 149 aa. unknown, similar
                     to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
                     MLCB33.03 (B2235_F3_140) (178 aa), fasta scores; opt: 498
                     z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2740"
                     /protein_id="CAA15536.1"
                     /db_xref="GI:2624262"
                     /db_xref="SPTREMBL:O33283"
                     /translation="MAELTETSPETPETTEAIRAVEAFLNALQNEDFDTVDAALGDDL
                     VYENVGFSRIRGGRRTATLLRRMQGRVGFEVKIHRIGADGAAVLTERTDALIIGPLRV
                     QFWVCGVFEVDDGRITLWRDYFDVYDMFKGLLRGLVALVVPSLKATL"
     gene            5355. .6932
                     /gene="PE_PGRS"
     CDS             5355. .6932
                     /gene="PE_PGRS"
                     /note="Rv2741, (MTV002.06), len: 525 aa; Member of M.
                     tuberculosis PE_ PGRS subfamily, similar to many eg.

Query Match             87.2%;  Score 21.8;  DB 1;  Length 56414;
Best Local Similarity   92.0%;  Pred. No. 1.6;
Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGA 25
          ||||| ||||||||||||||||||||

Db  38186 GAATTATCATCAGCAATCTTGCAGA 38210


RESULT   6
AC010703
LOCUS       AC010703   180588 bp    DNA              HTG       21-JAN-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-12N8, WORKING DRAFT SEQUENCE,
            6 unordered pieces.
ACCESSION   AC010703
VERSION     AC010703.2  GI:6466524
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180588)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180588)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Nov 23, 1999 this sequence version replaced gi:5903069.
             -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 651
                 Center clone name: RP11-12N8
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer Bodipy;    0% of reads
                 Chemistry: Dye-terminator;  100% of reads
                 Assembly program  Phrap; version 0.990319
                 Consensus quality: 173888 bases at least Q40
                 Consensus quality: 177848 bases at least Q30
                 Consensus quality: 179211 bases at least Q20
                 Insert size: 174844; agarose-fp
                 Insert size: 180338; sum-of-contigs
                 Quality coverage:  6.0 in Q20 bases; agarose-fp
                 Quality coverage:  5.8 in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3645: contig of 3645 bp in length
            *     3646     3695: gap of unknown length
            *     3696     9186: contig of 5491 bp in length
            *     9187     9236: gap of unknown length
            *     9237    29348: contig of 20112 bp in length
            *    29349    29398: gap of unknown length
            *    29399    57481: contig of 28083 bp in length
            *    57482    57531: gap of unknown length
            *    57532    99700: contig of 42169 bp in length
            *    99701    99750: gap of unknown length
            *    99751   180588: contig of 80838 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180588
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-12N8"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    54206 a  33718 c  34726 g  57686 t    252 others
ORIGIN


  Query Match             79.2%;  Score 19.8;  DB 45;  Length 180588;
  Best Local Similarity   91.3%;  Pred. No. 18;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 ATTCCCATCAGCAATCTTGCAGA 25
           ||||||||||||||| || ||||
Db  115526 ATTCCCATCAGCAATGTTTCAGA 115548


RESULT   7
AC018616
LOCUS       AC018616   165737 bp    DNA             HTG       28-DEC-1999
DEFINITION  Homo sapiens chromosome 8 clone RP11-157I4 map 8, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC018616
VERSION     AC018616.2  GI:6642692
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165737)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-157I4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165737)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
            O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
            Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
            and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 28, 1999 this sequence version replaced gi:6573859.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5176
                Center clone name: 157_I_4
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 162119 bases at least Q40
                Consensus quality: 164141 bases at least Q30
                Consensus quality: 164860 bases at least Q20
                Insert size: 164000; agarose-fp
                Insert size: 165737; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3324: contig of 3324 bp in length
            *                   gap of unknown length
            *     3325    15022: contig of 11698 bp in length
            *                   gap of unknown length
            *    15023    59126: contig of 44104 bp in length
            *                   gap of unknown length
            *    59127   165737: contig of 106611 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165737
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-157I4"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    52701 a  31191 c  29989 g  51853 t      3 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 44;  Length 165737;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           ||||| |||||| ||||||||||
Db   29366 GAATTGCCATCAAGAATCTTGCAG 29389


RESULT   8
AC011958/c
LOCUS       AC011958   172513 bp    DNA             HTG       17-OCT-1999
DEFINITION  Homo sapiens chromosome 18 clone 223_G_18 map 18, *** SEQUENCING IN
            PROGRESS ***, 14 unordered pieces.
ACCESSION   AC011958
VERSION     AC011958.1  GI:6065416
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172513)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone 223_G_18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172513)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    52204: contig of 52204 bp in length
            *                    gap of unknown length
            *    52205    54784: contig of 2580 bp in length
            *                    gap of unknown length
            *    54785    57254: contig of 2470 bp in length
            *                    gap of unknown length
            *    57255    61619: contig of 4365 bp in length
            *                    gap of unknown length
            *    61620    65444: contig of 3825 bp in length
            *                    gap of unknown length
            *    65445    69687: contig of 4243 bp in length
            *                    gap of unknown length
            *    69688    74201: contig of 4514 bp in length
            *                    gap of unknown length
            *    74202    82818: contig of 8617 bp in length
            *                    gap of unknown length
            *    82819    93900: contig of 11082 bp in length
            *                    gap of unknown length
            *    93901   106828: contig of 12928 bp in length
            *                    gap of unknown length
            *   106829   121552: contig of 14724 bp in length
            *                    gap of unknown length
            *   121553   139340: contig of 17788 bp in length
            *                    gap of unknown length
            *   139341   167314: contig of 27974 bp in length
            *                    gap of unknown length
            *   167315   172513: contig of 5199 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172513
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="223_G_18"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    52777 a  31276 c  30717 g  56518 t   1225 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 42;  Length 172513;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           || |||||| ||| ||||||||||
Db   93323 GACTTCCCAGCAGAAATCTTGCAG 93300


RESULT   9
AC020694
LOCUS       AC020694   176147 bp    DNA             HTG       25-JAN-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-637N15 map 17, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC020694
VERSION     AC020694.3  GI:6751731
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176147)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-637N15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176147)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenester,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 25, 2000 this sequence version replaced gi:6730831.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5275
                Center clone name: 637_N_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 169002 bases at least Q40
                Consensus quality: 173024 bases at least Q30
                Consensus quality: 174690 bases at least Q20
                Insert size: 181000; agarose-fp
                Insert size: 176147; sum-of-contigs
                Quality coverage: 5.5 in Q20 bases; agarose-fp
                Quality coverage: 5.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1341: contig of 1341 bp in length
            *                     gap of unknown length
            *     1342     2832: contig of 1491 bp in length
            *                     gap of unknown length
            *     2833     4501: contig of 1669 bp in length
            *                     gap of unknown length
            *     4502     6602: contig of 2101 bp in length
            *                     gap of unknown length
            *     6603     9487: contig of 2885 bp in length
            *                     gap of unknown length
            *     9488    12146: contig of 2659 bp in length
            *                     gap of unknown length
            *    12147    17027: contig of 4881 bp in length
            *                     gap of unknown length
            *    17028    18738: contig of 1711 bp in length
            *                     gap of unknown length
```

```
            *    18739    23050: contig of 4312 bp in length
            *                    gap of unknown length
            *    23051    29398: contig of 6348 bp in length
            *                    gap of unknown length
            *    29399    36930: contig of 7532 bp in length
            *                    gap of unknown length
            *    36931    43762: contig of 6832 bp in length
            *                    gap of unknown length
            *    43763    51759: contig of 7997 bp in length
            *                    gap of unknown length
            *    51760    60023: contig of 8264 bp in length
            *                    gap of unknown length
            *    60024    71480: contig of 11457 bp in length
            *                    gap of unknown length
            *    71481    86002: contig of 14522 bp in length
            *                    gap of unknown length
            *    86003   102872: contig of 16870 bp in length
            *                    gap of unknown length
            *   102873   119707: contig of 16835 bp in length
            *                    gap of unknown length
            *   119708   138179: contig of 18472 bp in length
            *                    gap of unknown length
            *   138180   158374: contig of 20195 bp in length
            *                    gap of unknown length
            *   158375   176147: contig of 17773 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176147
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-637N15"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    52653 a  37885 c  37479 g  48129 t      1 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 45;  Length 176147;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAG 24
           ||| | |||||||||||||| |||
Db   24580 GAAATGCCATCAGCAATCTTACAG 24603


RESULT  10
AL135797
LOCUS       AL135797   189642 bp    DNA              HTG       23-DEC-1999
DEFINITION  Homo sapiens chromosome 1 clone GS1-77J13, *** SEQUENCING IN
            PROGRESS ***, 24 unordered pieces.
ACCESSION   AL135797
VERSION     AL135797.2  GI:6981935
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189642)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 16, 2000 this sequence version replaced gi:6634922.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00195  Length: 7518bp
            Contig_ID: 00457  Length: 1121bp
Contig_ID: 00494  Length: 1050bp
Contig_ID: 00497  Length: 15915bp
Contig_ID: 00567  Length: 1001bp
Contig_ID: 00584  Length: 1493bp
Contig_ID: 00759  Length: 15040bp
Contig_ID: 01075  Length: 21731bp
Contig_ID: 01114  Length: 1447bp
Contig_ID: 01130  Length: 1284bp
Contig_ID: 01211  Length: 13125bp
Contig_ID: 01214  Length: 5959bp
Contig_ID: 01223  Length: 7513bp
Contig_ID: 01356  Length: 6435bp
Contig_ID: 01439  Length: 1068bp
Contig_ID: 01455  Length: 4961bp
Contig_ID: 01653  Length: 7152bp
Contig_ID: 01657  Length: 2705bp
Contig_ID: 01772  Length: 6392bp
Contig_ID: 01786  Length: 17050bp
Contig_ID: 01839  Length: 7643bp
Contig_ID: 02043  Length: 12585bp
Contig_ID: 02075  Length: 3089bp
Contig_ID: 02204  Length: 7965bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     7518: contig of 7518 bp in length
*     7519 8318: gap of      800 bp
*     8319     9439: contig of 1121 bp in length
*     9440 10239: gap of      800 bp
*    10240    11289: contig of 1050 bp in length
*    11290 12089: gap of      800 bp
*    12090    28004: contig of 15915 bp in length
*    28005 28804: gap of      800 bp
*    28805    29805: contig of 1001 bp in length
*    29806 30605: gap of      800 bp
*    30606    32098: contig of 1493 bp in length
*    32099 32898: gap of      800 bp
*    32899    47938: contig of 15040 bp in length
*    47939 48738: gap of      800 bp
*    48739    70469: contig of 21731 bp in length
*    70470 71269: gap of      800 bp
*    71270    72716: contig of 1447 bp in length
*    72717 73516: gap of      800 bp
*    73517    74800: contig of 1284 bp in length
*    74801 75600: gap of      800 bp
*    75601    88725: contig of 13125 bp in length
*    88726 89525: gap of      800 bp
*    89526    95484: contig of 5959 bp in length
*    95485 96284: gap of      800 bp
*    96285   103797: contig of 7513 bp in length
*   103798 104597: gap of      800 bp
*   104598   111032: contig of 6435 bp in length
*   111033 111832: gap of      800 bp
*   111833   112900: contig of 1068 bp in length
*   112901 113700: gap of      800 bp
*   113701   118661: contig of 4961 bp in length
*   118662 119461: gap of      800 bp
*   119462   126613: contig of 7152 bp in length
*   126614 127413: gap of      800 bp
*   127414   130118: contig of 2705 bp in length
*   130119 130918: gap of      800 bp
*   130919   137310: contig of 6392 bp in length
*   137311 138110: gap of      800 bp
*   138111   155160: contig of 17050 bp in length
*   155161 155960: gap of      800 bp
*   155961   163603  contig of 7643 bp in length
*   163604 164403: gap of      800 bp
*   164404   176988: contig of 12585 bp in length
```

```
            *   176989 177788: gap of      800 bp
            *   177789   180877: contig of 3089 bp in length
            *   180878 181677: gap of      800 bp
            *   181678   189642: contig of 7965 bp in length.
FEATURES              Location/Qualifiers
     source           1. .189642
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /clone="GS1-77J13"
                      /clone_lib="Genome_Systems_ReleaseI"
BASE COUNT    54211 a  32267 c  32632 g  52130 t  18402 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 32;  Length 189642;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAGA 25
           ||||||||  || ||||||||||||
Db    3443 AATTCCCAGGAGAAATCTTGCAGA 3466


RESULT  11
AL133553/c
LOCUS       AL133553   198436 bp    DNA             HTG       01-FEB-2000
DEFINITION  Homo sapiens chromosome 1 clone GS1-174L6, *** SEQUENCING IN
            PROGRESS ***, 11 unordered pieces.
ACCESSION   AL133553
VERSION     AL133553.2  GI:6982865
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198436)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 16, 2000 this sequence version replaced gi:6634914.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00211  Length: 1051bp
            Contig_ID: 00591  Length: 30642bp
            Contig_ID: 00683  Length: 28153bp
            Contig_ID: 01036  Length: 2595bp
            Contig_ID: 01176  Length: 1621bp
            Contig_ID: 01269  Length: 30486bp
            Contig_ID: 01360  Length: 1103bp
            Contig_ID: 01683  Length: 20866bp
            Contig_ID: 01965  Length: 34873bp
            Contig_ID: 02118  Length: 11466bp
            Contig_ID: 02309  Length: 27580bp.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1051: contig of 1051 bp in length
            *     1052 1851: gap of      800 bp
            *     1852    32493: contig of 30642 bp in length
            *    32494 33293: gap of      800 bp
            *    33294    61446: contig of 28153 bp in length
            *    61447 62246: gap of      800 bp
            *    62247    64841: contig of 2595 bp in length
            *    64842 65641: gap of      800 bp
            *    65642    67262: contig of 1621 bp in length
            *    67263 68062: gap of      800 bp
            *    68063    98548: contig of 30486 bp in length
            *    98549 99348: gap of      800 bp
            *    99349   100451: contig of 1103 bp in length
            *   100452 101251: gap of      800 bp
            *   101252   122117: contig of 20866 bp in length
            *   122118 122917: gap of      800 bp
            *   122918   157790: contig of 34873 bp in length
            *   157791 158590: gap of      800 bp
            *   158591   170056: contig of 11466 bp in length
            *   170057 170856: gap of      800 bp
            *   170857   198436: contig of 27580 bp in length.
FEATURES              Location/Qualifiers
     source           1. .198436
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /clone="GS1-174L6"
                      /clone_lib="Genome_Systems_ReleaseI"
BASE COUNT    59085 a  37013 c  35907 g  58401 t   8030 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 33;  Length 198436;
  Best Local Similarity   87.5%;  Pred. No. 36;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 AATTCCCATCAGCAATCTTGCAGA 25
           ||||||||  || ||||||||||||
Db   19128 AATTCCCAGGAGAAATCTTGCAGA 19105


RESULT  12
OAKRT211/c
LOCUS       OAKRT211     1401 bp    DNA             MAM       24-FEB-1999
DEFINITION  Ovis aries gene for hair keratin type II intermediate filament
            protein, N-terminus.
ACCESSION   X73455
VERSION     X73455.1  GI:313238
KEYWORDS    keratin intermediate filament.
SOURCE      sheep.
  ORGANISM  Ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Artiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
            Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1401)
  AUTHORS   Powell,B.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-1993) B.C. Powell, University of Adelaide, Dept
            of Biochemistry, Adelaide, South Australia 5001, AUSTRALIA
REFERENCE   2  (bases 1 to 1401)
  AUTHORS   Mojica,F.J., Juez,G. and Rodriguez-Valera,F.
  TITLE     Transcription at different salinities of Haloferax mediterranei
            sequences adjacent to partially modified PstI sites
  JOURNAL   Mol. Microbiol. 9 (3), 613-621 (1993)
  MEDLINE   94018655
FEATURES              Location/Qualifiers
     source           1. .1401
                      /organism="Ovis aries"
                      /db_xref="taxon:9940"
                      /clone_lib="genomic cosmid"
     misc_signal      442. .448
                      /note="AP1 box"
     misc_signal      527. .534
                      /note="HK1 box"
     CAAT_signal      560. .572
     CAAT_signal      636. .648
     TATA_signal      692. .699
     exon             <794. .1162
```

```
                     /gene="KRT2.11"
                     /number=1
     gene            794. .1401
                     /gene="KRT2.11"
     CDS             794. .>1162
                     /gene="KRT2.11"
                     /codon_start=1
                     /product="hair keratin type II intermediate filament"
                     /protein_id="CAA51834.1"
                     /db_xref="GI:313239"
                     /db_xref="SPTREMBL:Q08126"
                     /translation="MTCGSYRAVPAFSCVSACGPRPGRCCITAAPYRGISCYRGLTGG
                     FGSRSVCGGFRAGSCGRSFGYRSGGVGGLSPPCITTVSVNESLLTPLNLEIDPNAQCV
                     KQEEKEQIKCLNNRFAAFIDK"
     intron          1163. .>1401
                     /gene="KRT2.11"
                     /number=1
BASE COUNT      332 a    384 c    401 g    284 t
ORIGIN


  Query Match           74.4%;  Score 18.6;  DB 3;  Length 1401;
  Best Local Similarity  84.0%;  Pred. No. 58;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || ||||||||| ||||||  ||||
Db     140 GATTTCCCATCAACAATCTCACAGA 116


RESULT  13
CPHYPAB
LOCUS       CPHYPAB      2468 bp    DNA             BCT       19-MAY-1999
DEFINITION  Chlamydophila caviae hypA and hypB genes for stress response
            proteins.
ACCESSION   X51404 M25101
VERSION     X51404.1  GI:40577
KEYWORDS    antigen; hypA gene; hypB gene; stress protein.
SOURCE      Chlamydophila caviae.
  ORGANISM  Chlamydophila caviae
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (bases 1 to 2468)
  AUTHORS   Morrison,R.P., Belland,R.J., Lyng,K. and Caldwell,H.D.
  TITLE     Chlamydial disease pathogenesis. The 57-kD chlamydial
            hypersensitivity antigen is a stress response protein
  JOURNAL   J. Exp. Med. 170 (4), 1271-1283 (1989)
  MEDLINE   90010784
REFERENCE   2  (bases 1 to 2468)
  AUTHORS   Morrison,R.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1989) to the EMBL/GenBank/DDBJ databases
COMMENT     hypB encodes the 57 kD chlamydial hypersensitivity antigen.
FEATURES             Location/Qualifiers
     source          1. .2468
                     /organism="Chlamydophila caviae"
                     /strain="GPIC"
                     /specific_host="Guinea Pig"
                     /db_xref="taxon:83557"
     gene            314. .622
                     /gene="hypA"
     CDS             314. .622
                     /gene="hypA"
                     /function="stress response protein"
                     /note="11.2 kD protein (AA 1-102)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypA protein"
                     /protein_id="CAA35765.1"
                     /db_xref="GI:40578"
                     /db_xref="SWISS-PROT:P15598"
                     /translation="MSDQATTLRIKPLGDRILVKREEEDSTARGGIILPDTAKKKQDR
                     AEVLVLGTGKRDKDGNVLPFEVTVGDTVLIDKYAGQELTVDGEEYVIVQESEVMAVLK
                     *
     gene            673. .2307
                     /gene="hypB"
     CDS             673. .2307
                     /gene="hypB"
                     /function="stress response protein"
                     /note="58.1 kD (AA 1-544)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypB protein"
                     /protein_id="CAA35766.1"
                     /db_xref="GI:40579"
                     /db_xref="SWISS-PROT:P15599"
                     /translation="MAAKNIKYNEDARKKIHKGVKTLAEAVKVTLGPKGRHVVIDKSF
                     GSPQVTKDGVTVAKEIELEDKHENMGAQMVKEVASKTADKAGDGTTTATVLAEAIYSE
                     GLRNVTAGANPMDLKRGIDKAVKVVVDEIKKISKPVQHHKEIAQVATISANNDAEIGN
                     LIAEAMEKVGKNGSITVEEAKGFETVLDVVEGMNFNRGYLSSYFSTNPETQECVLEEA
                     LVLIYDKKISGIKDFLPVLQQVAESGRPLLIIAEDIEGEALATLVVNRLRAGFRVCAV
                     KAPGFGDRRKAMLEDIAILTGGQLISEELGMKLENTTLAMLGKAKKVIVSKEDTTIVE
                     GLGSKEDIESRCESIKKQIEDSTSDYDKEKLQERLAKLSGGVAVIRVGAATEIEMKEK
                     KDRVDDAQNATLAAVEEGILPGGGTALVRCIPTLEAFIPILTNEDEQIGARIVLKALS
                     APLKQIAANAGKEGAIICQQVLSRSSSEGYDALRDAYTDMIEAGILDPTKVTRCALES
                     AASVAGLLLTTEALIADIPEEKSSSAPAMPGAGMDY"
BASE COUNT      855 a    462 c    514 g    637 t
ORIGIN


  Query Match           74.4%;  Score 18.6;  DB 1;  Length 2468;
  Best Local Similarity  84.0%;  Pred. No. 60;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || ||||||| || ||||||| |||
Db      66 GACTTCCCATTAGAAATCTTGAAGA 90


RESULT  14
AC013732/c
LOCUS       AC013732   141602 bp    DNA             HTG       01-FEB-2000
DEFINITION  Homo sapiens clone RP11-550O23, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION   AC013732
VERSION     AC013732.3  GI:6850499
KEYWORDS    HTG; HTGS_PHASE1; HTG  _DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141602)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141602)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 1, 2000 this sequence version replaced gi:6447131.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0550023
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 82%
            Sequencing vector: plasmid; 18%
            Chemistry: Dye-primer ET; 82% of reads
            Chemistry: Dye-terminator Big Dye; 18% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 135904 bases at least Q40
            Consensus quality: 138143 bases at least Q30
```

```
          Consensus quality: 139347 bases at least Q20
          Insert size: 158000; agarose-fp
          Insert size: 141602; sum-of-contigs
          Quality coverage: 4.27 in Q20 bases; agarose-fp
          Quality coverage: 4.43 in Q20 bases; sum-of-contigs
          ----------------------------------------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *         1     2730: contig of 2730 bp in length
          *                     gap of unknown length
          *      2731     6729: contig of 3999 bp in length
          *                     gap of unknown length
          *      6730    12002: contig of 5273 bp in length
          *                     gap of unknown length
          *     12003    17495: contig of 5493 bp in length
          *                     gap of unknown length
          *     17496    24522: contig of 7027 bp in length
          *                     gap of unknown length
          *     24523    34424: contig of 9902 bp in length
          *                     gap of unknown length
          *     34425    45429: contig of 11005 bp in length
          *                     gap of unknown length
          *     45430    58769: contig of 13340 bp in length
          *                     gap of unknown length
          *     58770    72377: contig of 13608 bp in length
          *                     gap of unknown length
          *     72378    92783: contig of 20406 bp in length
          *                     gap of unknown length
          *     92784   110713: contig of 17930 bp in length
          *                     gap of unknown length
          *    110714   141602: contig of 30889 bp in length.
FEATURES             Location/Qualifiers
     source          1. .141602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-550O23"
BASE COUNT    43027 a  29405 c  28598 g  40561 t     11 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 52;  Length 141602;
  Best Local Similarity   84.0%;  Pred. No. 72;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || ||| || |||||| ||||||||
Db   48086 GATTTCACAACAGCAACCTTGCAGA 48062


RESULT  15
HSJ831A12/c
LOCUS       HSJ831A12  172175 bp    DNA              HTG       16-FEB-2000
DEFINITION  Homo sapiens chromosome 1 clone RP5-831A12 map q24.1-25.3, ***
            SEQUENCING IN PROGRESS ***, 21 unordered pieces.
ACCESSION   AL122019
VERSION     AL122019.12  GI:6996474
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172175)
  AUTHORS   Thomas,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 17, 2000 this sequence version replaced gi:6983423.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.  Work on the sequence is in
            progress and the release of this data is based on the understanding
            that the sequence may change as work continues.  The sequence may
            be contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Contig_ID: 00111  Length: 1226bp
            Contig_ID: 00263  Length: 1683bp
            Contig_ID: 00324  Length: 1593bp
            Contig_ID: 00399  Length: 1724bp
            Contig_ID: 00439  Length: 1357bp
            Contig_ID: 00595  Length: 1204bp
            Contig_ID: 00654  Length: 1180bp
            Contig_ID: 00741  Length: 1045bp
            Contig_ID: 00781  Length: 12878bp
            Contig_ID: 00914  Length: 2743bp
            Contig_ID: 00957  Length: 1614bp
            Contig_ID: 01011  Length: 30320bp
            Contig_ID: 01059  Length: 1258bp
            Contig_ID: 01179  Length: 1018bp
            Contig_ID: 01220  Length: 1579bp
            Contig_ID: 01349  Length: 1226bp
            Contig_ID: 01381  Length: 1622bp
            Contig_ID: 01492  Length: 1219bp
            Contig_ID: 01542  Length: 87043bp
            Contig_ID: 01787  Length: 1486bp
            Contig_ID: 01825  Length: 1157bp.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1226: contig of 1226 bp in length
            *    1227  2026: gap of      800 bp
            *    2027     3709: contig of 1683 bp in length
            *    3710  4509: gap of      800 bp
            *    4510     6102: contig of 1593 bp in length
            *    6103  6902: gap of      800 bp
            *    6903     8626: contig of 1724 bp in length
            *    8627  9426: gap of      800 bp
            *    9427    10783: contig of 1357 bp in length
            *   10784 11583: gap of      800 bp
            *   11584    12787: contig of 1204 bp in length
            *   12788 13587: gap of      800 bp
            *   13588    14767: contig of 1180 bp in length
            *   14768 15567: gap of      800 bp
            *   15568    16612: contig of 1045 bp in length
            *   16613 17412: gap of      800 bp
            *   17413    30290: contig of 12878 bp in length
            *   30291 31090: gap of      800 bp
            *   31091    33833: contig of 2743 bp in length
            *   33834 34633: gap of      800 bp
            *   34634    36247: contig of 1614 bp in length
            *   36248 37047: gap of      800 bp
            *   37048    67367: contig of 30320 bp in length
            *   67368 68167: gap of      800 bp
            *   68168    69425: contig of 1258 bp in length
            *   69426 70225: gap of      800 bp
            *   70226    71243: contig of 1018 bp in length
            *   71244 72043: gap of      800 bp
            *   72044    73622: contig of 1579 bp in length
            *   73623 74422: gap of      800 bp
            *   74423    75648: contig of 1226 bp in length
            *   75649 76448: gap of      800 bp
            *   76449    78070: contig of 1622 bp in length
            *   78071 78870: gap of      800 bp
            *   78871    80089: contig of 1219 bp in length
            *   80090 80889: gap of      800 bp
            *   80890   167932: contig of 87043 bp in length
```

```
           *   167933 168732: gap of      800 bp
           *   168733   170218: contig of 1486 bp in length
           *   170219 171018: gap of      800 bp
           *   171019   172175: contig of 1157 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172175
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24.1-25.3"
                     /clone="RP5-831A12"
                     /clone_lib="RPCI-5"
BASE COUNT    43936 a  30479 c  31749 g  49966 t  16045 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 33;  Length 172175;
  Best Local Similarity   84.0%;  Pred. No. 73;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGA 25
           || ||| || ||| ||||||||||||
Db  122299 GACTTCTCAGCAGAAATCTTGCAGA 122275


Search completed: June 22, 2000, 14:58:31
Job time: 17926 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:16 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-13
Perfect score: 18
Sequence: 1 GAGACCAAAAACAACGAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	16.4	91.1	389	35	AA550152	AA550152 1265m3 gm
c	2	16.4	91.1	425	39	AA840739	AA840739 CFR21 Fru
	3	16.4	91.1	477	37	AA734816	AA734816 vp38c05.r
	4	16	88.9	346	82	CNS00YSV	AL096589 Arabidops
	5	16	88.9	435	82	CNS00SNE	AL088616 Arabidops
	6	16	88.9	521	82	CNS00US0	AL091374 Arabidops
	7	16	88.9	1009	82	CNS005YQ	AL061969 Drosophil
	8	15.4	85.6	168	79	AW237496	AW237496 xm72f05.x
	9	15.4	85.6	175	27	AA002265	AA002265 zh82d02.s
c	10	15.4	85.6	199	70	AW152312	AW152312 xg44e07.x
	11	15.4	85.6	229	39	AA829559	AA829559 od97a05.s
c	12	15.4	85.6	235	38	AA808572	AA808572 ob38a12.s
c	13	15.4	85.6	239	59	AI771403	AI771403 EST252503
c	14	15.4	85.6	252	72	AV287500	AV287500 AV287500
c	15	15.4	85.6	286	27	C04637	C04637 C04637 Huma
	16	15.4	85.6	300	62	AV199183	AV199183 AV199183
	17	15.4	85.6	303	24	H86827	H86827 yt07d05.s1
	18	15.4	85.6	307	60	AV165505	AV165505 AV165505
	19	15.4	85.6	313	44	AI263087	AI263087 qz35a08.x
	20	15.4	85.6	315	21	F08843	F08843 HSC2TE102 n
	21	15.4	85.6	331	27	AA025469	AA025469 ze86h01.s
	22	15.4	85.6	332	23	R91155	R91155 yp94f01.s1
	23	15.4	85.6	338	43	AI204102	AI204102 qd73d06.x
	24	15.4	85.6	345	45	AI369467	AI369467 ta53g01.x
	25	15.4	85.6	349	47	AI538639	AI538639 tp70e12.x
	26	15.4	85.6	351	24	H86255	H86255 yt05d05.s1
c	27	15.4	85.6	352	50	F26315	F26315 HSPD13753 H
	28	15.4	85.6	361	47	AI472758	AI472758 ta13e06.x
	29	15.4	85.6	361	79	AW244110	AW244110 xa51a10.x
	30	15.4	85.6	363	35	AA552117	AA552117 nk05g02.s
	31	15.4	85.6	363	47	AI500631	AI500631 tn98g10.x
c	32	15.4	85.6	366	37	AA678107	AA678107 zi14g10.s
	33	15.4	85.6	367	79	AW271979	AW271979 xs12e04.x
c	34	15.4	85.6	373	27	C02102	C02102 HUMGS000642
	35	15.4	85.6	379	24	H96074	H96074 yv98c09.s1
	36	15.4	85.6	384	45	AI367968	AI367968 qv93e10.x
	37	15.4	85.6	386	39	AA878539	AA878539 oj19e08.s
	38	15.4	85.6	387	22	R68290	R68290 yi06h08.s1
	39	15.4	85.6	387	27	AA045693	AA045693 zk81d10.s
	40	15.4	85.6	388	105	AQ305338	AQ305338 HS_2037_B
	41	15.4	85.6	389	37	AA715096	AA715096 nx94f04.s
c	42	15.4	85.6	389	50	F36793	F36793 HSPD34690 H
	43	15.4	85.6	396	33	AA449755	AA449755 zx07f08.s
	44	15.4	85.6	399	40	AA918276	AA918276 on42d02.s
	45	15.4	85.6	401	107	AQ443475	AQ443475 GSSTc0017

ALIGNMENTS

RESULT 1
AA550152/c
LOCUS AA550152 389 bp DNA EST 11-AUG-1997
DEFINITION 1265m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic clone 1265m, mRNA sequence.
ACCESSION AA550152
VERSION AA550152.1 GI:2320404
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 389)
AUTHORS Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z., Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N., Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A., Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su,X.-Z., Thompson,J.K., Vital,F., Wellems,T.E. and Werner,E.
TITLE Current status of the Plasmodium falciparum genome project
JOURNAL Mol. Biochem. Parasitol. 79, 1-12 (1996)
MEDLINE 97001675
COMMENT On Nov 29, 1993 this sequence version replaced gi:636032.
Contact: Dame JB
Department of Pathobiology
University of Florida
2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: dame@icbr.ifas.ufl.edu
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .389
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="1265m"
/clone_lib="gmbPfHB3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBlueScript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were polished using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBlueScript SK(+). Recombinant plasmids transformed E. coli XL1-Blue."
BASE COUNT 147 a 37 c 35 g 168 t 2 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 35; Length 389;
Best Local Similarity 94.4%; Pred. No. 7.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
|| ||||||||||||||||
Db 339 GAAACCAAAAACAACGAA 322

RESULT 2
AA840739/c
LOCUS AA840739 425 bp mRNA EST 02-MAR-1998
DEFINITION CFR21 Fruit cDNA library of Hot pepper Capsicum annuum cDNA clone CFR21 5' similar to transcription factor BAF1 homolog (P33293), mRNA sequence.
ACCESSION AA840739
VERSION AA840739.1 GI:2922075
KEYWORDS EST.
SOURCE Capsicum annuum.
ORGANISM Capsicum annuum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
REFERENCE 1 (bases 1 to 425)
AUTHORS Hong,S.T., Chung,J.E., An,G. and Kim,S.R.
TITLE Analysis of 176 expressed sequence tags generated from cDNA clones of Hot pepper by single-pass sequencing
JOURNAL Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153529.
Contact: Kim SR
Laboratory of Plant Molecular Biology, Department of Life Science
Sogang University
C.P.O Box 1142, Seoul 100-611, Korea
Tel: 02 705 8456
Fax: 02 704 3601
Email: Sungkim@ccs.sogang.ac.kr
histidine-glutamine rich protein
Insert Length: 1 Std Error: 0.00
Seq primer: T3 primer
High quality sequence stop: 400.
FEATURES Location/Qualifiers
source 1. .425
/organism="Capsicum annuum"
/cultivar="Happy Dry"

/db_xref="taxon:4072"
/clone="CFR21"
/clone_lib="Fruit cDNA library of Hot pepper"
/sex="hermaphrodite"
/dev_stage="green fruit with 0.5 to 2 centimeter in length"
/note="Organ: Fruit; Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: EcoRI; Oligo(dT) primed cDNA from poly(A) RNA. In vivo excised to pBluescript plasmid."

BASE COUNT 96 a 138 c 69 g 122 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 39; Length 425;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
|| ||||||||||||||||
Db 204 GAAACCAAAAACAACGAA 187

RESULT 3
AA734816
LOCUS AA734816 477 bp mRNA EST 07-JAN-1998
DEFINITION vp38c05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA clone IMAGE:1078952 5', mRNA sequence.
ACCESSION AA734816
VERSION AA734816.1 GI:2756483
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 477)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 9, 1995 this sequence version replaced gi:803097.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:593248
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 458.
FEATURES Location/Qualifiers
source 1. .477
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1078952"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. "
BASE COUNT 126 a 107 c 111 g 133 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 37; Length 477;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
|||||||||||||||| ||
Db 175 GAGACCAAAAACAACAAA 192

RESULT 4
CNS00YSV
LOCUS CNS00YSV 346 bp DNA GSS 29-JUN-1999
DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC T17F6 of TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.
ACCESSION AL096589
VERSION AL096589.1 GI:5304744
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 346)
AUTHORS Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 346)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
FEATURES Locatio. Qualifiers
source 1. .346
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T17F6"
/note="end : T7"
BASE COUNT 128 a 57 c 103 g 58 t
ORIGIN

Query Match 88.9%; Score 16; DB 82; Length 346;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACG 16
||||||||||||||||
Db 231 GAGACCAAAAACAACG 246

RESULT 5
CNS00SNE
LOCUS CNS00SNE 438 bp DNA GSS 29-JUN-1999
DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC T2E24 of TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.
ACCESSION AL088616
VERSION AL088616.1 GI:5289756
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core

```
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 435)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="TAMU"
                     /clone="T2E24"
                     /note="end : SP6"
BASE COUNT      150 a     79 c    122 g     84 t
ORIGIN


  Query Match             88.9%;  Score 16;  DB 82;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACG 16
          ||||||||||||||||
Db    234 GAGACCAAAAACAACG 249


RESULT   6
CNS00US0
LOCUS       CNS00US0      521 bp    DNA             GSS       28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence SP6 end of BAC T7M20 of
            TAMU library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL091374
VERSION     AL091374.1  GI:5292528
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 521)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="TAMU"
                     /clone="T7M20"
                     /note="end : SP6"
BASE COUNT      184 a     96 c    144 g     97 t
ORIGIN


  Query Match             88.9%;  Score 16;  DB 82;  Length 521;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACG 16
          ||||||||||||||||
Db    345 GAGACCAAAAACAACG 360


RESULT   7
CNS005YQ
LOCUS       CNS005YQ      1009 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T3 end of BAC #
            BACR12J24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061969
VERSION     AL061969.1  GI:4943765
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1009)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1009
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR12J24"
                     /note="end : TET3"
BASE COUNT      232 a     302 c    261 g     198 t     16 others
ORIGIN


  Query Match             88.9%;  Score 16;  DB 82;  Length 1009;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 GACCAAAAACAACGAA 18
          ||||||||||||||||
Db     10 GACCAAAAACAACGAA 25


RESULT   8
AW237496
LOCUS       AW237496      168 bp    mRNA            EST       13-DEC-1999
DEFINITION  xm72f05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2689761 3',
            mRNA sequence.
ACCESSION   AW237496
VERSION     AW237496.1  GI:6569885
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 168)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135427.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2689761"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT       61 a     41 c     23 g     43 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 79;  Length 168;
  Best Local Similarity   94.1%;  Pred. No. 2.2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGA 17
          ||||||||||||||| |
Db    111 GAGACCAAAAACAACCA 127


RESULT  9
AA002265
LOCUS       AA002265       175 bp    mRNA            EST       07-MAY-1997
DEFINITION  zh82d02.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:427779 3', mRNA sequence.
ACCESSION   AA002265
VERSION     AA002265.1  GI:1445180
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 21, 1992 this sequence version replaced gi:279399.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 573   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 124.
FEATURES             Location/Qualifiers
     source          1. .175
                     /organism="Homo sapiens"
                     /db_xref="GDB:1327548"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:427779"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.   1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       59 a     40 c     24 g     48 t      4 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 27;  Length 175;
  Best Local Similarity   94.1%;  Pred. No. 2.2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGA 17
          ||||||||||||||| |
Db     94 GAGACCAAAAACAACCA 110


RESULT  10
AW152312/c
LOCUS       AW152312       199 bp    mRNA            EST       03-NOV-1999
DEFINITION  xg44e07.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2630436 3'
            similar to TR:O75197 O75197 LIPOPROTEIN RECEPTOR RELATED PROTEIN 5.
            ;, mRNA sequence.
ACCESSION   AW152312
VERSION     AW152312.1  GI:6200212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5407048.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -4[illegible]UP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .199
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2630436"
/clone_lib="NCI_CGAP_Ut1"
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 45 a 49 c 60 g 44 t [illegible]1 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 70; Length 199;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 37 GAGACCAACAACAACGA 21

RESULT 11
AA829559
LOCUS AA829559 229 bp mRNA EST 20-FEB-1998
DEFINITION od97a05.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1384304, mRNA sequence.
ACCESSION AA829559
VERSION AA829559.1 GI:2902658
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 229)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150810.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.

FEATURES Location/Qualifiers
source 1. .229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1384304"
/clone_lib="NCI_CGAP_Ov2"
/sex="female"
/tissue_type="ovary"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive ovarian tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 87 a 45 c 45 g 52 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 39; Length 229;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
| |||||||||||||||
Db 144 ACACCAAAAACAACGAA 160

RESULT 12
AA808572/c
LOCUS AA808572 235 bp mRNA EST 19-FEB-1998
DEFINITION ob38a12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1333630 3' similar to TR:Q90706 Q90706 CLE7. ;, mRNA sequence.
ACCESSION AA808572
VERSION AA808572.1 GI:2877978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 235)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153343.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 1061 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .235
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1333630"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 55 a 55 c 54 g 71 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 38; Length 235;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
||||||||| |||||||
Db 84 AGACCAAAATCAACGAA 68

RESULT 13
AI771403/c
LOCUS AI771403 239 bp mRNA EST 29-JUN-1999
DEFINITION EST252503 tomato ovary, TAMU Lycopersicon esculentum cDNA clone cLED29F17, mRNA sequence.
ACCESSION AI771403
VERSION AI771403.1 GI:5269444
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 239)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948524.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES Location/Qualifiers
source 1. .239
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED29F17"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lamda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT 74 a 48 c 47 g 70 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 59; Length 239;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
| ||||||||||||||||
Db 200 ACACCAAAAACAACGAA 184

RESULT 14
AV287500/c
LOCUS AV287500 252 bp mRNA EST 09-NOV-1999
DEFINITION AV287500 RIKEN full-length enriched, 11 days pregnant adult female ovary and uterus Mus musculus cDNA clone 5031436013, mRNA sequence.
ACCESSION AV287500
VERSION AV287500.1 GI:6298004
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 252)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189328.
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .252
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5031436013"
/clone_lib="RIKEN full-length enriched, 11 days pregnant adult female ovary and uterus"
/sex="female"
/tissue_type="ovary and uterus"
/dev_stage="11 days pregnant, adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of

sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 77 a 47 c 53 g 75 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 72; Length 252;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 17 GAGACCAAGAACAACGA 1

RESULT 15
C04637/c
LOCUS C04637 286 bp mRNA EST 30-JUL-1996
DEFINITION C04637 Human heart cDNA (YNakamura) Homo sapiens cDNA clone 3NHC3725, mRNA sequence.
ACCESSION C04637
VERSION C04637.1 GI:1467888
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 286)
AUTHORS Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and Nakamura,Y.
TITLE Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing
JOURNAL Genomics 35 (1), 231-235 (1996)
MEDLINE 96299762
COMMENT Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES Location/Qualifiers
source 1. .286
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="3NHC3725"
/clone_lib="Human heart cDNA (YNakamura)"
/dev_stage="adult"
/note="Organ: heart; normalized directionally cloned cDNA from adult heart"
BASE COUNT 96 a 42 c 54 g 92 t 2 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 27; Length 286;
Best Local Similarity 94.1%; Pred. No. 2.1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||||||||| |
Db 51 GAGACCAAAAACAACCA 35

Search completed: June 22, 2000, 12:08:20
Job time: 10414 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:36 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-13
Perfect score: 18
Sequence: 1 GAGACCAAAAACAACGAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.8	82.2	874	2	US-08-120-607A-10	Sequence 10, Appl
	2	14.8	82.2	4138	2	US-08-447-411-75	Sequence 75, Appl
	3	14.8	82.2	4138	4	US-08-662-227-33	Sequence 33, Appl
	4	14.8	82.2	5211	2	US-08-447-411-1	Sequence 1, Appli
	5	14.4	80.0	5194	3	US-08-642-846-1	Sequence 1, Appli
	6	14.4	80.0	6755	5	US-08-931-999-4	Sequence 4, Appli
	7	13.8	76.7	974	1	US-08-220-606B-29	Sequence 29, Appl
	8	13.8	76.7	1095	1	US-08-220-606B-1	Sequence 1, Appli
c	9	13.8	76.7	1743	1	US-08-171-382-3	Sequence 3, Appli
c	10	13.8	76.7	1743	1	US-08-309-420-3	Sequence 3, Appli
c	11	13.8	76.7	1743	1	US-08-309-419-3	Sequence 3, Appli
c	12	13.8	76.7	1743	6	PCT-US95-11856-3	Sequence 3, Appli
c	13	13.8	76.7	1743	6	PCT-US95-11878-3	Sequence 3, Appli
c	14	13.8	76.7	2392	1	US-08-171-382-5	Sequence 5, Appli
c	15	13.8	76.7	2392	1	US-08-309-420-5	Sequence 5, Appli
c	16	13.8	76.7	2392	1	US-08-309-419-5	Sequence 5, Appli
c	17	13.8	76.7	2392	6	PCT-US95-11856-5	Sequence 5, Appli
c	18	13.8	76.7	2392	6	PCT-US95-11878-5	Sequence 5, Appli
	19	13.8	76.7	2621	4	US-08-553-619B-8	Sequence 8, Appli
c	20	13.8	76.7	4379	2	US-08-592-214A-17	Sequence 17, Appl
c	21	13.8	76.7	7032	3	US-08-149-097D-24	Sequence 24, Appl
c	22	13.8	76.7	7655	2	US-08-619-554-1	Sequence 1, Appli
	23	13.4	74.4	420	6	PCT-US94-07297-34	Sequence 34, Appl
c	24	13.4	74.4	619	4	US-08-966-316-1	Sequence 1, Appli
	25	13.4	74.4	2839	6	PCT-US94-07297-38	Sequence 38, Appl
c	26	13.4	74.4	2912	1	US-07-931-943-1	Sequence 1, Appli
c	27	13.4	74.4	2912	3	US-08-624-601-1	Sequence 1, Appli
c	28	13.4	74.4	2915	1	US-07-931-943-4	Sequence 4, Appli
c	29	13.4	74.4	2915	3	US-08-624-601-4	Sequence 4, Appli
	30	13.4	74.4	3048	1	US-08-188-228-47	Sequence 47, Appl
	31	13.4	74.4	3048	1	US-08-332-643-41	Sequence 41, Appl
	32	13.4	74.4	3048	1	US-08-332-638-47	Sequence 47, Appl
	33	13.4	74.4	3478	1	US-08-396-479B-1	Sequence 1, Appli
	34	13.4	74.4	3478	2	US-08-818-823-1	Sequence 1, Appli
	35	13.4	74.4	19011	1	US-08-310-356-36	Sequence 36, Appl
	36	13.4	74.4	19557	6	PCT-US92-06300-1	Sequence 1, Appli
c	37	13.2	73.3	69	6	PCT-US92-01015-12	Sequence 12, Appl
c	38	13.2	73.3	72	6	PCT-US92-01015-3	Sequence 3, Appli
c	39	13.2	73.3	81	1	US-08-026-145-3	Sequence 3, Appli
c	40	13.2	73.3	96	1	US-07-872-673B-4	Sequence 4, Appli
c	41	13.2	73.3	134	1	US-08-090-193-8	Sequence 8, Appli
c	42	13.2	73.3	134	3	US-08-488-031-8	Sequence 8, Appli
c	43	13.2	73.3	134	3	US-08-486-569-8	Sequence 8, Appli
c	44	13.2	73.3	134	3	US-08-488-027-8	Sequence 8, Appli
c	45	13.2	73.3	134	5	US-08-482-658-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT   1
US-08-120-607A-10/c
; Sequence 10, Application US/08120607A
; Patent No. 5762939
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Gale E.
;    APPLICANT:  Volvovitz, Franklin
;    APPLICANT:  Wilkinson, Bethanie
;    APPLICANT:  Hackett, Craig
;    TITLE OF INVENTION:  A METHOD FOR PRODUCING INFLUENZA
;    TITLE OF INVENTION:  HEMAGGLUTININ MULTIVALENT VACCINES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas J. Kowalski
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/120,507A
;      FILING DATE:  September 13, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kowalski, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  674501-2001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  874 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Influenza virus
US-08-120-607A-10

  Query Match           82.2%;  Score 14.8;  DB 2;  Length 874;
```

```
  Best Local Similarity   88.9%;  Pred. No. 89;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGAA 18
           ||||||| |||||| |||
Db     431 GAGACCAGAAACAAGGAA 414


RESULT   2
US-08-447-411-75
; Sequence 75, Application US/08447411
; Patent No. 5773243
;  GENERAL INFORMATION:
;    APPLICANT:  FRITZINGER, DAVID C.
;    APPLICANT:  BREDEHORST, REINHARD
;    APPLICANT:  VOGEL, CARL-WILHELM
;    TITLE OF INVENTION:  DNA ENCODING COBRA C3, CVF1, AND CVF2
;    NUMBER OF SEQUENCES:  81
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,411
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/043,747
;      FILING DATE:  07-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773243man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-101-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  75:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4138 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..4001
US-08-447-411-75


  Query Match             82.2%;  Score 14.8;  DB 2;  Length 4138;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGAA 18
           |||||||||  |||||||
Db    2961 GAGACCAAACTCAACGAA 2978


RESULT   3
US-08-662-227-33
; Sequence 33, Application US/08662227
; Patent No. 5922320
;  GENERAL INFORMATION:
;    APPLICANT:  VOGEL, CARL-WILHELM
;    APPLICANT:  BREDEHORST, REINHORST
;    APPLICANT:  KOCK, MICHAEL
;    APPLICANT:  FRITZINGER, DAVID
;    TITLE OF INVENTION:  RECOMBINANT PROCVF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/662,227
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-0107-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4138 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-662-227-33


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 4138;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGAA 18
           |||||||||  |||||||
Db    2961 GAGACCAAACTCAACGAA 2978


RESULT   4
US-08-447-411-1
; Sequence 1, Application US/08447411
; Patent No. 5773243
;  GENERAL INFORMATION:
;    APPLICANT:  FRITZINGER, DAVID C.
;    APPLICANT:  BREDEHORST, REINHARD
;    APPLICANT:  VOGEL, CARL-WILHELM
;    TITLE OF INVENTION:  DNA ENCODING COBRA C3, CVF1, AND CVF2
;    NUMBER OF SEQUENCES:  81
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/447,411
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/043,747
;      FILING DATE:  07-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773243man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-101-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5211 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..4961
US-08-447-411-1


  Query Match             82.2%;  Score 14.8;  DB 2;  Length 5211;
  Best Local Similarity   88.9%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGAA 18
           |||||||||  |||||||
Db    3936 GAGACCAAACTCAACGAA 3953


RESULT   5
US-08-642-846-1
; Sequence 1, Application US/08642846
; Patent No. 5886151
;  GENERAL INFORMATION:
;    APPLICANT:  HOSTETTER, MARGARET K.
;    APPLICANT:  GALE, CHERYL A.
;    APPLICANT:  BENDEL, CATHERINE M.
;    APPLICANT:  TAO, NIAN-JUN
;    APPLICANT:  KENDRICK, KATHLEEN
;    TITLE OF INVENTION:  CANDIDA ALBICANS GENE, INTEGRIN-LIKE
;    TITLE OF INVENTION:  PROTEIN, ANTIBODIES, AND METHODS OF USE
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
;      STREET:  119 NORTH FOURTH STREET, SUITE 203
;      CITY:  MINNEAPOLIS
;      STATE:  MINNESOTA
;      COUNTRY:  USA
;      ZIP:  55401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/642,846
;      FILING DATE:  03-MAY-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MUETING, ANN M.
;      REGISTRATION NUMBER:  33,977
;      REFERENCE/DOCKET NUMBER:  110.00280101
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-305-1217
;      TELEFAX:  612-305-1228
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5194 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-642-846-1


  Query Match             80.0%;  Score 14.4;  DB 3;  Length 5194;
  Best Local Similarity   93.8%;  Pred. No. 1.6e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GACCAAAAACAACGAA 18
           || ||||||||||||||
Db     831 GAACAAAAACAACGAA 846


RESULT   6
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
;  GENERAL INFORMATION:
;    APPLICANT:  Iandolo, John J.
;    APPLICANT:  Crupper, Scott S.
;    TITLE OF INVENTION:  Broad Spectrum Chemotherapeutic Peptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hovey, Williams, Timmons & Collins
;      STREET:  2405 Grand Boulevard, Suite 400
;      CITY:  Kansas City
;      STATE:  Missouri
;      COUNTRY:  U.S.A.
;      ZIP:  64108
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,999
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/710,561
;      FILING DATE:  19-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Collins, John M.
;      REGISTRATION NUMBER:  26,262
;      REFERENCE/DOCKET NUMBER:  25043-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  816/474-9050
;      TELEFAX:  816/474-9057
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6755 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  UT0007
US-08-931-999-4


  Query Match             80.0%;  Score 14.4;  DB 5;  Length 6755;
  Best Local Similarity   93.8%;  Pred. No. 1.6e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGA 17
```

```
          ||||||||||||||| |
Db    6189 AGACCAAAAACAACAA 6204


RESULT   7
US-08-220-606B-29
; Sequence 29, Application US/08220606B
; Patent No. 5641661
;  GENERAL INFORMATION:
;    APPLICANT:  Kumagai, Monto H.
;    APPLICANT:  Genadi, Sverlow J.
;    TITLE OF INVENTION:  Pichia Pastoris Alcohol Oxidase ZZA1 and
;    TITLE OF INVENTION:  ZZA2 Regulatory Regions for Heterologous Gene Expression
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/220,606B
;      FILING DATE:  25-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P.
;      REGISTRATION NUMBER:  25,227
;      REFERENCE/DOCKET NUMBER:  8129-065
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-3660
;      TELEFAX:  415-854-3694
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  974 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-220-606B-29


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 974;
  Best Local Similarity   88.2%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||| |||||||||||| ||
Db     898 AGAAGAAAAACAACTAA 914


RESULT   8
US-08-220-606B-1
; Sequence 1, Application US/08220606B
; Patent No. 5641661
;  GENERAL INFORMATION:
;    APPLICANT:  Kumagai, Monto H.
;    APPLICANT:  Genadi, Sverlow J.
;    TITLE OF INVENTION:  Pichia Pastoris Alcohol Oxidase ZZA1 and
;    TITLE OF INVENTION:  ZZA2 Regulatory Regions for Heterologous Gene Expression
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/220,606B
;      FILING DATE:  25-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P.
;      REGISTRATION NUMBER:  25,227
;      REFERENCE/DOCKET NUMBER:  8129-065
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-3660
;      TELEFAX:  415-854-3694
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1095 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-220-606B-1


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 1095;
  Best Local Similarity   88.2%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||| |||||||||||| ||
Db    1019 AGAACAAAAACAACTAA 1035


RESULT   9
US-08-171-382-3/c
; Sequence 3, Application US/08171382
; Patent No. 5472856
;  GENERAL INFORMATION:
;    APPLICANT:  Harris, Crafford A.
;    APPLICANT:  Goldstein, Gideon
;    APPLICANT:  Siekierka, John J.
;    APPLICANT:  Talle, Mary Anne
;    APPLICANT:  Shenbagamurthi, Ponniah
;    APPLICANT:  Culler, Michael D.
;    APPLICANT:  Setcavage, Diane R.
;    TITLE OF INVENTION:  Recombinant Human Thymopoietin Proteins
;    TITLE OF INVENTION:  and Uses Therefor
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/171,382
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  IRI43USA
;    TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 238..1599
US-08-171-382-3

Query Match 76.7%; Score 13.8; DB 1; Length 1743;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
||||||||||||| ||
Db 1528 AGACCAAAAACAAAAAA 1512

RESULT 10
US-08-309-420-3/c
; Sequence 3, Application US/08309420
; Patent No. 5591588
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; APPLICANT: Culler, Michael
; TITLE OF INVENTION: Method for the Diagnosis of Depression
; TITLE OF INVENTION: Based on Monitoring Blood Levels of Arginine Vasopressin
; TITLE OF INVENTION: and/or Thymopoietin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 No. 5591588ristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,420
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI46USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9207
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 238..1599
US-08-309-420-3

Query Match 76.7%; Score 13.8; DB 1; Length 1743;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
||||||||||||| ||
Db 1528 AGACCAAAAACAAAAAA 1512

RESULT 11
US-08-309-419-3/c
; Sequence 3, Application US/08309419
; Patent No. 5593842
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; APPLICANT: Culler, Michael
; TITLE OF INVENTION: Method of Measuring Thymopoietin
; TITLE OF INVENTION: Proteins in Plasma and Serum
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 No. 5593842ristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,419
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI45BUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 238..1599
US-08-309-419-3

Query Match 76.7%; Score 13.8; DB 1; Length 1743;
Best Local Similarity 88.2%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
||||||||||||| ||
Db 1528 AGACCAAAAACAAAAAA 1512

RESULT 12
PCT-US95-11856-3/c
; Sequence 3, Application PC/TUS9511856
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Method of Measuring
; TITLE OF INVENTION: Thymopoietin Proteins in Plasma and Serum
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 Norristown Road, Box 457

```
;      CITY:  Spring House
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version#1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11856
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/309,419
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  IRI45BPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 540-9206
;      TELEFAX:  (215) 540-5818
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1743 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  238..1599
PCT-US95-11856-3


  Query Match             76.7%;  Score 13.8;  DB 6;  Length 1743;
  Best Local Similarity   88.2%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||||||||||||||  ||
Db    1528 AGACCAAAAACAAAAAA 1512


RESULT  13
PCT-US95-11878-3/c
; Sequence 3, Application PC/TUS9511878
;  GENERAL INFORMATION:
;    APPLICANT:  Immunobiology Research, Institute Inc.
;    TITLE OF INVENTION:  Method for the Diagnosis of
;    TITLE OF INVENTION:  Depression Based on Monitoring Blood Levels of
;    TITLE OF INVENTION:  Arginine Vasopressin and/or Thymopoietin
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  321 Norristown Road, Box 457
;      CITY:  Spring House
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version#1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/309,420
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  IRI46PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 540-9206
;      TELEFAX:  (215) 540-5818
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1743 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  238..1599
PCT-US95-11878-3


  Query Match             76.7%;  Score 13.8;  DB 6;  Length 1743;
  Best Local Similarity   88.2%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||||||||||||||  ||
Db    1528 AGACCAAAAACAAAAAA 1512


RESULT  14
US-08-171-382-5/c
; Sequence 5, Application US/08171382
; Patent No. 5472856
;  GENERAL INFORMATION:
;    APPLICANT:  Harris, Crafford A.
;    APPLICANT:  Goldstein, Gideon
;    APPLICANT:  Siekierka, John J.
;    APPLICANT:  Talle, Mary Anne
;    APPLICANT:  Shenbagamurthi, Ponniah
;    APPLICANT:  Culler, Michael D.
;    APPLICANT:  Setcavage, Diane R.
;    TITLE OF INVENTION:  Recombinant Human Thymopoietin Proteins
;    TITLE OF INVENTION:  and Uses Therefor
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/171,382
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  IRI43USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  241..1275
US-08-171-382-5


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 2392;
  Best Local Similarity   88.2%;  Pred. No. 2.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||||||||||||  ||
Db    1204 AGACCAAAAACAAAAAA 1188


RESULT  15
US-08-309-420-5/c
; Sequence 5, Application US/08309420
; Patent No. 5591588
;  GENERAL INFORMATION:
;    APPLICANT:  Goldstein, Gideon
;    APPLICANT:  Culler, Michael
;    TITLE OF INVENTION:  Method for the Diagnosis of Depression
;    TITLE OF INVENTION:  Based on Monitoring Blood Levels of Arginine Vasopressin
;    TITLE OF INVENTION:  and/or Thymopoietin
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  321 No. 5591588ristown Road, Box 457
;      CITY:  Spring House
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1..5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/309,420
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  IRI46USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 540-9207
;      TELEFAX:  (215) 540-5818
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  241..1275
US-08-309-420-5


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 2392;
  Best Local Similarity   88.2%;  Pred. No. 2.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 AGACCAAAAACAACGAA 18
           ||||||||||||  ||
Db    1204 AGACCAAAAACAAAAAA 1188
```

Search completed: June 22, 2000, 15:21:38
Job time: 18776 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:40 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-13
Perfect score: 18
Sequence: 1 GAGACCAAAAACAACGAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	1	V49514	Mycobacterium sp.
	2	18	100.0	18	1	V49612	AlaDH derived olig
	3	18	100.0	1245	1	V49626	Mycobacterium tube
	4	18	100.0	1260	1	V49510	Mycobacterium sp.
	5	18	100.0	1260	1	V49625	Mycobacterium tube
	6	15.4	85.6	1038	1	X51740	DNA encoding a hum
c	7	15.4	85.6	2213	1	X13016	Enterococcus faeca
	8	15.4	85.6	4351	1	V70397	LRP5 isoform 2 lon
	9	15.4	85.6	4843	1	V70395	LRP5 cDNA longest
	10	15.4	85.6	4915	1	V70398	LRP5 isoform 3 put
	11	15.4	85.6	5022	1	V85549	LRP5 isoform 5 cDN
	12	15.4	85.6	5098	1	V70396	LRP5 isoform 1 cDN
	13	15.4	85.6	5125	1	V86019	Lrp5 isoform 3 cDN
	14	15.4	85.6	5162	1	V85550	LRP5 isoform 6 cDN
	15	15.4	85.6	5166	1	V85551	LRP5 isoform 2 cDN
	16	15.4	85.6	5263	1	V70400	LRP5 isoform 4 cDN
c	17	15	83.3	110000	1	T58840_0	Mycoplasma genital
c	18	14.8	82.2	874	1	V23092	5' end of the infl
	19	14.8	82.2	2304	1	X05857	Rat pheromone rece
	20	14.8	82.2	3584	1	X05818	Rat pheromone rece
	21	14.8	82.2	4138	1	Q77791	Cobra partial CVF2
	22	14.8	82.2	4607	1	X13231	Enterococcus faeca
	23	14.8	82.2	5211	1	Q77789	Pre-pro-cobra C3 c
c	24	14.8	82.2	10669	1	V52190	Streptococcus pneu
c	25	14.8	82.2	15363	1	X13216	Enterococcus faeca
	26	14.8	82.2	110000	1	T58840_2	Continuation (3 of
	27	14.4	80.0	5194	1	X25885	C.albicans alpha-I
	28	14.4	80.0	6755	1	V21511	Staphylococcal bac
	29	14.4	80.0	9797	1	X13487	Enterococcus faeca
	30	13.8	76.7	169	1	Q05209	Primer binding to
	31	13.8	76.7	213	1	T25729	Human gene signatu
c	32	13.8	76.7	316	1	V87092	EST clone BD209. N
c	33	13.8	76.7	327	1	X40630	Human secreted pro
	34	13.8	76.7	498	1	V90089	EST clone DB159. N
	35	13.8	76.7	548	1	Q06153	Circumsporozoite a
	36	13.8	76.7	560	1	Q06152	Circumsporozoite a
	37	13.8	76.7	587	1	T64830	Coral cobra venom
	38	13.8	76.7	763	1	V65204	DNA encoding high-
c	39	13.8	76.7	943	1	X13413	Enterococcus faeca
	40	13.8	76.7	1095	1	Q72574	P. pastoris alcoho
	41	13.8	76.7	1095	1	T80523	Partial alcohol ox
c	42	13.8	76.7	1743	1	Q92774	Human thymopoietin
c	43	13.8	76.7	1743	1	T13450	Human thymopoietin
c	44	13.8	76.7	1743	1	T26960	Thymopoietin-beta
c	45	13.8	76.7	2175	1	Q29268	Human calcium chan

ALIGNMENTS

RESULT 1
V49514
ID V49514 standard; DNA; 18 BP.
AC V49514;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F2.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 11 A; 4 C; 3 G; 0 U;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 1 GAGACCAAAAACAACGAA 18

RESULT 2
V49612
ID V49612 standard; DNA; 18 BP.
AC V49612;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F2.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 11 A; 4 C; 3 G; 0 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 1 GAGACCAAAAACAACGAA 18

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 97 GAGACCAAAAACAACGAA 114

RESULT 5
V49625
ID V49625 standard; DNA; 1250 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 97 GAGACCAAAAACAACGAA 114

RESULT 6
X51740
ID X51740 standard; DNA; 1038 BP.
AC X51740;
DT 17-JUN-1999 (first entry)
DE DNA encoding a human secreted protein.
KW Human secreted protein; cancer; immune disorder; infection;
KW inflammatory disorder; skin disorder; tumour; atherosclerosis;
KW restenosis; autoimmune disorder; Alzheimer's disease;
KW peripheral neuropathy; trauma; spinal cord injury; allergy;
KW hematopoietic disorder; skeletal disorder; neurological disorder;
KW arthritic disorder; asthma; immunodeficiency disease; AIDS;
KW transplant rejection; ss.
OS Homo sapiens.
PN WO9911293-A1.
PD 11-MAR-1999.
PF 03-SEP-1998; U18360.
PR 12-SEP-1997; US-058974.
PR 05-SEP-1997; US-057626.
PR 05-SEP-1997; US-057663.
PR 05-SEP-1997; US-057669.
PR 12-SEP-1997; US-058666.
PR 12-SEP-1997; US-058667.
PR 12-SEP-1997; US-058973.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Lafleur DW, Moore PA, Olsen HS,
PI Rosen GA, Ruben SM, Shi Y;
DR WPI; 99-204988/17.
DR P-PSDB; Y12953.
PT New isolated human genes and the secreted polypeptides they encode
PT - useful for diagnosis and treatment of e.g. neurological disorders,
PT tumours, immune disorders, inflammation or haematological disorders
PS Claim 1; Page 173; 215pp; English.
CC X51701-55 encode human secreted proteins. The polynucleotides and
CC their corresponding secreted polypeptides are useful for preventing,
CC treating or ameliorating medical conditions, e.g. by protein or gene
CC therapy. Pathological conditions can also be diagnosed by determining
CC the amount of the new polypeptides in a sample or by determining the
CC presence of mutations in the new polynucleotides. Specific uses are
CC described for each polynucleotide, based on which tissues they are
CC most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, immune disorders, infection,
CC inflammatory disorders, skin disorders, tumours, atherosclerosis,
CC restenosis, autoimmune disorders, Alzheimer's disease, peripheral
CC neuropathies, trauma, spinal cord injuries, allergy, hematopoietic
CC disorders, skeletal disorders, neurological disorders, arthritic
CC disorders, asthma, immunodeficiency diseases, AIDS and transplant
CC rejection. The polypeptides are also useful for identifying their
CC binding partners.
SQ Sequence 1038 BP; 327 A; 214 C; 248 G; 247 T;

Query Match 85.6%; Score 15.4; DB 1; Length 1038;
Best Local Similarity 94.1%; Pred. No. 98;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
|||||| ||||||||||
Db 188 GAGACCGAAAACAACGA 204

RESULT 7
X13016/c
ID X13016 standard; DNA; 2213 BP.
AC X13016;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:79.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 558-559; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 2213 BP; 721 A; 342 C; 427 G; 720 T;

Query Match 85.6%; Score 15.4; DB 1; Length 2213;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AGACCAAAAACAACGAA 18
|||||||||| ||||||
Db 2151 AGACCAAAAAAAACGAA 2135

RESULT 8
V70397
ID V70397 standard; cDNA; 4351 BP.
AC V70397;
DT 10-FEB-1999 (first entry)
DE LRP5 isoform 3 longest open reading frame (also isoform 4,5,6).
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.

PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 7; Fig 11b; 200pp; English.
CC The present sequence represents LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3) isoform 2 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 4351 BP; 875 A; 1435 C; 1309 G; 732 T;

Query Match 85.6%; Score 15.4; DB 1; Length 4351;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 1489 GAGACCAATAACAACGA 1505

RESULT 9
V70395
ID V70395 standard; cDNA; 4843 BP.
AC V70395;
DT 10-FEB-1999 (first entry)
DE LRP5 cDNA longest open reading frame.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Example 1; Fig 5b; 200pp; English.
CC The present sequence represents the longest open reading frame of LRP5
CC (low density lipoprotein (LDL) receptor related protein, previously
CC designated LRP-3) cDNA from the present invention. Nucleic acid
CC molecules (NAMs) encoding LRP5 can be used for determining if an
CC individual is susceptible to insulin dependent diabetes mellitus (IDDM).
CC The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 4843 BP; 953 A; 1601 C; 1478 G; 811 T;

Query Match 85.6%; Score 15.4; DB 1; Length 4843;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 1981 GAGACCAATAACAACGA 1997

RESULT 10
V70398
ID V70398 standard; cDNA; 4915 BP.
AC V70398;
DT 10-FEB-1999 (first entry)
DE LRP5 isoform 3 putative open reading frame.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Example 1; Fig 12c; 200pp; English.
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the putative open reading frame of LRP5 isoform 3.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 4915 BP; 992 A; 1614 C; 1474 G; 835 T;

Query Match 85.6%; Score 15.4; DB 1; Length 4915;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 2052 GAGACCAATAACAACGA 2068

RESULT 11
V85549
ID V85549 standard; cDNA; 5022 BP.
AC V85549;

DT 10-FEB-1999 (first entry)
DE LRP5 isoform 5 cDNA.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 7; Fig 14; 200pp; English.
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the LRP5 isoform 5 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin depe dent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 5022 BP; 1036 A; 1606 C; 1503 G; 877 T;

Query Match 85.6%; Score 15.4; DB 1; Length 5022;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 1951 GAGACCAATAACAACGA 1967

RESULT 12
V70396
ID V70396 standard; cDNA; 5098 BP.
AC V70396;
DT 10-FEB-1999 (first entry)
DE LRP5 isoform 1 cDNA.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Claim 3; Fig 5a; 200pp; English.
CC The present sequence represents LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3) isoform 1 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 5098 BP; 103 A; 1645 C; 1541 G; 877 T;

Query Match 85.6%; Score 15.4; DB 1; Length 5098;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGA 17
||||||||| |||||||||
Db 2029 GAGACCAATAACAACGA 2045

RESULT 13
V86019
ID V86019 standard; cDNA; 5125 BP.
AC V86019;
DT 10-FEB-1999 (first entry)
DE Lrp5 isoform 3 cDNA.
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS Homo sapiens.
PN WO9846743-A1.
PD 22-OCT-1998.
PF 15-APR-1998; G01102.
PR 05-JUN-1997; US-048740.
PR 15-APR-1997; US-043553.
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI Phillips MS, Todd JA, Twells RCJ;
DR WPI; 98-594573/50.
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
PS Example 1; Fig 12a; 200pp; English.
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the Lrp5 isoform 3 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 5125 BP; 1069 A; 1640 C; 1518 G; 898 T;

Query Match 85.6%; Score 15.4; DB 1; Length 5125;

```
  Best Local Similarity   94.1%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACG 17
           ||||||||  ||||||||
Db    2055 GAGACCAATAACAACGA 2071


RESULT  14
V85550
ID   V85550 standard; cDNA; 5162 BP.
AC   V85550;
DT   10-FEB-1999  (first entry)
DE   LRP5 isoform 6 cDNA.
KW   LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW   endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW   glomerulonephritis; inflammation; viral infection; osteoporosis;
KW   hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS   Homo sapiens.
PN   WO9846743-A1.
PD   22-OCT-1998.
PF   15-APR-1998; G01102.
PR   05-JUN-1997; US-048740.
PR   15-APR-1997; US-043553.
PA   (MERI ) MERCK & CO INC.
PA   (WELL ) WELLCOME TRUST LTD.
PI   Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI   Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI   Phillips MS, Todd JA, Twells RCJ;
DR   WPI; 98-594573/50.
PT   New isolated LDL-receptor related protein - used to develop products
PT   for treating, e.g. elevated triglyceride levels, diabetes,
PT   autoimmune disorders, inflammation or Alzheimer's disease
PS   Claim 7; Fig 15a; 200pp; English.
CC   The present invention describes LRP5 (low density lipoprotein (LDL)
CC   receptor related protein, previously designated LRP-3). The present
CC   sequence represents the LRP5 isoform 6 cDNA.
CC   Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC   if an individual is susceptible to insulin dependent diabetes mellitus
CC   (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC   in the serum of an individual. Therapies that affect LRP5 may also be
CC   useful in the treatment of autoimmune diseases such as
CC   glomerulonephritis, diseases and disorders involving disruption of
CC   endocytosis and/or antigen presentation, cytokine clearance and/or
CC   inflammation, viral infection, pathogenic bacterial toxin contamination,
CC   elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC   osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC   from the present invention can also be used for detection, diagnosis and
CC   drug screening.
SQ   Sequence   5162 BP;    1062 A;    1665 C;    1527 G;    908 T;


  Query Match             85.6%;  Score 15.4;  DB 1;  Length 5162;
  Best Local Similarity   94.1%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGA 17
           |||||||| ||||||||
Db    2092 GAGACCAATAACAACGA 2108


RESULT  15
V85551
ID   V85551 standard; cDNA; 5166 BP.
AC   V85551;
DT   10-FEB-1999  (first entry)
DE   LRP5 isoform 2 cDNA.
KW   LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW   endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW   glomerulonephritis; inflammation; viral infection; osteoporosis;
KW   hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
OS   Homo sapiens.
PN   WO9846743-A1.
PD   22-OCT-1998.
PF   15-APR-1998; G01102.
PR   05-JUN-1997; US-048740.
PR   15-APR-1997; US-043553.
PA   (MERI ) MERCK & CO INC.
PA   (WELL ) WELLCOME TRUST LTD.
PI   Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW,
PI   Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y,
PI   Phillips MS, Todd JA, Twells RCJ;
DR   WPI; 98-594573/50.
PT   New isolated LDL-receptor related protein - used to develop products
PT   for treating, e.g. elevated triglyceride levels, diabetes,
PT   autoimmune disorders, inflammation or Alzheimer's disease
PS   Claim 8; Fig 11a; 200pp; English.
CC   The present invention describes LRP5 (low density lipoprotein (LDL)
CC   receptor related protein, previously designated LRP-3). The present
CC   sequence represents the LRP5 isoform 2 cDNA.
CC   Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC   if an individual is susceptible to insulin dependent diabetes mellitus
CC   (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC   in the serum of an individual. Therapies that affect LRP5 may also be
CC   useful in the treatment of autoimmune diseases such as
CC   glomerulonephritis, diseases and disorders involving disruption of
CC   endocytosis and/or antigen presentation, cytokine clearance and/or
CC   inflammation, viral infection, pathogenic bacterial toxin contamination,
CC   elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC   osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC   from the present invention can also be used for detection, diagnosis and
CC   drug screening.
SQ   Sequence   5166 BP;    1082 A;    1654 C;    1530 G;    899 T;


  Query Match             85.6%;  Score 15.4;  DB 1;  Length 5166;
  Best Local Similarity   94.1%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GAGACCAAAAACAACGA 17
           |||||||| ||||||||
Db    2106 GAGACCAATAACAACGA 2122


Search completed: June 22, 2000, 15:12:43
Job time: 18089 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:57:58 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-13
Perfect score: 18
Sequence: 1 GAGACCAAAAACAACGAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87616	A87616 Sequence 13
2		18	100.0	720	5	A89759	A89759 Sequence 16
3		18	100.0	1194	2	MTU92472	U92472 Mycobacteri
4		18	100.0	1208	5	A89752	A89752 Sequence 9
5		18	100.0	1209	5	A87612	A87612 Sequence 9
6		18	100.0	1228	5	A87608	A87608 Sequence 5
7		18	100.0	1228	5	A89748	A89748 Sequence 5
8		18	100.0	1229	5	A87610	A87610 Sequence 7
9		18	100.0	1229	5	A89750	A89750 Sequence 7
10		18	100.0	1235	5	A87606	A87606 Sequence 3
11		18	100.0	1235	5	A87609	A87609 Sequence 6
12		18	100.0	1235	5	A87611	A87611 Sequence 8
13		18	100.0	1235	5	A89746	A89746 Sequence 3
14		18	100.0	1235	5	A89749	A89749 Sequence 6
15		18	100.0	1235	5	A89751	A89751 Sequence 8
16		18	100.0	1236	5	A87613	A87613 Sequence 10
17		18	100.0	1236	5	A89753	A89753 Sequence 10
18		18	100.0	1237	5	A87607	A87607 Sequence 4
19		18	100.0	1237	5	A89747	A89747 Sequence 4
20		18	100.0	1245	5	A87605	A87605 Sequence 2
21		18	100.0	1245	5	A89745	A89745 Sequence 2
22		18	100.0	1260	5	A87604	A87604 Sequence 1
23		18	100.0	1260	5	A89744	A89744 Sequence 1
24		18	100.0	2584	1	MTALADH	X63069 M.tuberculo
25		18	100.0	56414	1	MTV002	AL008967 Mycobacte
c 26		17	94.4	125113	33	AL136234	AL136234 Homo sapi
c 27		17	94.4	159515	12	AC002327	AC002327 Mus muscu
28		16.4	91.1	10246	12	MMU251787	AJ251787 Mus muscu
c 29		16.4	91.1	21011	34	CELF12A10	U28731 Caenorhabdi
c 30		16.4	91.1	33706	34	CEF28D9	Z81518 Caenorhabdi
31		16.4	91.1	35350	34	CEC30H6	Z81044 Caenorhabdi
c 32		16.4	91.1	40572	34	CELC14F11	U39645 Caenorhabdi
33		16.4	91.1	44019	10	HSU160A4	Z80900 Human DNA s
34		16.4	91.1	57799	55	AC023379	AC023379 Homo sapi
c 35		16.4	91.1	83379	32	DMBR43C24	AL133496 Drosophil
36		16.4	91.1	95604	11	AC004100	AC004100 Homo sapi
37		16.4	91.1	110000	32	CEY116A8_5	Continuation (6 of
c 38		16.4	91.1	110000	32	CEY53H1_1	Continuation (2 of
c 39		16.4	91.1	201352	33	AL133271	AL133271 Homo sapi
c 40		16.4	91.1	233941	32	CEY52B11	Z98867 Caenorhabdi
c 41		16.4	91.1	253307	34	PFMAL3P7	AL034559 Plasmodiu
c 42		16	88.9	69034	35	AF007261	AF007261 Reclinomo
43		16	88.9	114056	50	AC019013	AC019013 Genomic S
44		16	88.9	264875	44	AC008440	AC008440 Homo sapi
c 45		15.4	85.6	705	2	AF124499	AF124499 Streptoco

ALIGNMENTS

```
RESULT   1
A87616
LOCUS       A87616          18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 13 from Patent WO9836089.
ACCESSION   A87616
VERSION     A87616.1  GI:6736256
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       11 a      4 c      3 g      0 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 29;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db      1 GAGACCAAAAACAACGAA 18


RESULT   2
A89759
LOCUS       A89759          720 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9832862.
ACCESSION   A89759
VERSION     A89759.1  GI:6738291
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      133 a    238 c    226 g    123 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 720;
  Best Local Similarity  100.0%;  Pred. No. 38;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db      9 GAGACCAAAAACAACGAA 26


RESULT   3
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity  100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT   4
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 5
A87612
LOCUS A87612 1209 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 9 from Patent WO9836089.
ACCESSION A87612
VERSION A87612.1 GI:6736252
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1209
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1209;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 6
A87608
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:6736248
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 7
A89748
LOCUS A89748 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9832862.
ACCESSION A89748
VERSION A89748.1 GI:6738282
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 8
A87610
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGACCAAAAACAACGAA 18
||||||||||||||||||
Db 82 GAGACCAAAAACAACGAA 99

RESULT 9
A89750
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9832862.

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  10
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  11
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE)  SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  12
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  13
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  14
A89749
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99


RESULT  15
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCAAAAACAACGAA 18
          ||||||||||||||||||
Db     82 GAGACCAAAAACAACGAA 99
```

Search completed: June 22, 2000, 14:58:00
Job time: 17895 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:13 ; Search time 5541.94 Seconds
(without alignments)
14.627 Million cell updates/sec

Title: US-09-362-485-12
Perfect score: 20
Sequence: 1 GCGCGTCGGTATTCCGACCG 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query %

	No.	Score	Match	Length	DB	ID	Description
c	1	16.4	82.0	691	84	B20846	B20846 T2903-T7 TA
c	2	16	80.0	836	102	AQ163541	AQ163541 nbxb0007O
	3	15.8	79.0	480	81	AJ282503	AJ282503 4A3A-AAY-
	4	15.8	79.0	625	85	AQ680935	AQ680935 HS_5478_B
	5	15.8	79.0	636	48	AI557542	AI557542 pt2.1-06.
c	6	15.4	77.0	212	34	C25459	C25459 C25459 Rice
c	7	15.4	77.0	268	20	D40400	D40400 RICS2352A R
c	8	15.4	77.0	356	34	C25152	C25152 C25152 Rice
c	9	15.4	77.0	570	49	AU057524	AU057524 AU057524
c	10	15.4	77.0	595	83	FR0043767	AL131259 Fugu rubr
	11	15.2	76.0	450	109	AQ648034	AQ648034 RPCI93-Dp
c	12	15.2	76.0	495	90	AQ808619	AQ808619 HS_5023_A
c	13	15.2	76.0	595	107	AQ448926	AQ448926 mgxb0022K
	14	15.2	76.0	609	51	AU070274	AU070274 AU070274
	15	15.2	76.0	665	49	AU057624	AU057624 AU057624
c	16	15.2	76.0	731	92	AQ363438	AQ363438 nbxb0059O
c	17	15.2	76.0	757	70	AW155406	AW155406 mgie0014A
c	18	14.8	74.0	239	20	D41228	D41228 RICS3581A R
c	19	14.8	74.0	353	34	C25348	C25348 C25348 Rice
c	20	14.8	74.0	539	61	AU075821	AU075821 AU075821
c	21	14.8	74.0	542	80	AU082152	AU082152 AU082152
c	22	14.8	74.0	567	80	AU082153	AU082153 AU082153
c	23	14.8	74.0	607	61	AU075822	AU075822 AU075822
	24	14.8	74.0	619	82	FR0006260	Z90070 F.rubripes
c	25	14.8	74.0	638	64	AW065578	AW065578 614060D05
c	26	14.8	74.0	682	80	AU082159	AU082159 AU082159
c	27	14.8	74.0	826	85	AQ740216	AQ740216 HS_5505_A
c	28	14.4	72.0	414	92	AQ363071	AQ363071 nbxb0053L
c	29	14.4	72.0	423	85	AQ686265	AQ686265 nbxb0068G
c	30	14.4	72.0	424	46	AI399024	AI399024 NCW10B4T3
c	31	14.4	72.0	509	107	AQ510127	AQ510127 nbxb0094H
	32	14.4	72.0	532	29	AA131737	AA131737 z135a07.s
c	33	14.4	72.0	563	102	AQ157539	AQ157539 nbxb0009L
c	34	14.4	72.0	606	105	AQ328030	AQ328030 nbxb0042C
c	35	14.4	72.0	621	106	AQ396422	AQ396422 mgxb0016H
c	36	14.4	72.0	693	105	AQ329482	AQ329482 nbxb0045G
	37	14.4	72.0	701	34	AA528350	AA528350 ne83d11.s
c	38	14.4	72.0	759	92	AQ573485	AQ573485 nbxb0081F
c	39	14.4	72.0	761	108	AQ577144	AQ577144 nbxb0090B
c	40	14.4	72.0	781	92	AQ573484	AQ573484 nbxb0081F
c	41	14.4	72.0	784	105	AQ327447	AQ327447 nbxb0040P
	42	14.2	71.0	335	102	AQ152429	AQ152429 HS_3108_A
	43	14.2	71.0	339	42	AI137143	AI137143 UI-R-C2p-
	44	14.2	71.0	345	34	AA534115	AA534115 nj73h03.s
	45	14.2	71.0	380	42	AI112300	AI112300 UI-R-Y0-m

ALIGNMENTS

RESULT 1
B20846/c
LOCUS B20846 691 bp DNA GSS 16-SEP-1997
DEFINITION T2903-T7 TAMU Arabidopsis thaliana genomic clone T2903, genomic survey sequence.
ACCESSION B20846
VERSION B20846.1 GI:2395900
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 691)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: T2903-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 69
High quality sequence stop: 151.
FEATURES Location/Qualifiers
source 1. .691
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T2903"
/clone_lib="TAMU"
/sex="hermaphrodite"
/note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII; Produced by Rod Wing"
BASE COUNT 83 a 311 c 58 g 236 t 3 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 84; Length 691;
Best Local Similarity 94.4%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGAC 18
|||| ||||||||||||
Db 106 GCGCATCGGTATTCCGAC 89

RESULT 2
AQ163541/c
LOCUS AQ163541 836 bp DNA GSS 12-SEP-1998
DEFINITION nbxb0007O19r CUGI Rice BAC Library Oryza sativa genomic clone nbxb0007O19r, genomic survey sequence.
ACCESSION AQ163541
VERSION AQ163541.1 GI:3553167
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 836)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 226.
FEATURES Location/Qualifiers
source 1. .836
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbxb0007O19r"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the

world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 174 a 186 c 199 g 275 t 2 others
ORIGIN

Query Match 80.0%; Score 16; DB 102; Length 836;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CGCGTCGGTATTCCGA 17
||||||||||||||||
Db 258 CGCGTCGGTATTCCGA 243

RESULT 3
AJ282503
LOCUS AJ282503 480 bp mRNA EST 07-FEB-2000
DEFINITION 4A3A-AAY-F-11-R Anopheles gambiae immune competent 4A3A Anopheles gambiae cDNA clone 4A3A-AAY-F-11, mRNA sequence.
ACCESSION AJ282503
VERSION AJ282503.1 GI:6930382
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anopheles.
REFERENCE 1 (bases 1 to 480)
AUTHORS Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.
TITLE Anopheles gambiae pilot gene discovery project: Identification of novel mosquito innate immunity genes from ESTs generated from immune competent cell lines
JOURNAL Unpublished (2000)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5409836.
Contact: Dimopoulos G
Fotis C. Kafatos laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES Location/Qualifiers
source 1. .480
/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-AAY-F-11"
/clone_lib="Anopheles gambiae immune competent 4A3A"
/cell_line="immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT 120 a 125 c 144 g 91 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 81; Length 480;
Best Local Similarity 89.5%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACC 19
|||||||||||||| || ||
Db 408 GCGCGTCGGTATTACGCCC 426

RESULT 4
AQ680935
LOCUS AQ680935 625 bp DNA GSS 25-JUN-1999
DEFINITION HS_5478_B1_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1054 Col=11 Row=D, genomic survey sequence.
ACCESSION AQ680935
VERSION AQ680935.1 GI:5229739
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 625)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT On Mar 23, 1999 this sequence version replaced gi:3325331.
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1054 row: D column: 11
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 625.
FEATURES Location/Qualifiers
source 1. .625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1054 Col=11 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 153 a 179 c 157 g 100 t 36 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 85; Length 625;
Best Local Similarity 89.5%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGCGTCGGTATTCCGACCG 20
|||||||||||||| || ||
Db 253 CGCGTCGGTATTCTGAACG 271

RESULT 5

```
AI557542
LOCUS       AI557542      636 bp    mRNA            EST       09-AUG-1999
DEFINITION  pt2.1-06.D12b.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION   AI557542
VERSION     AI557542.1  GI:4489905
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Huang,G.M., Ng,W.l., Farkas,J., He,L., Liang,H.A., Gordon,D., Yu,J.
            and Hood,L.
  TITLE     Prostate cancer expression profiling by cDNA sequencing analysis
  JOURNAL   Genomics 59 (2), 178-186 (1999)
  MEDLINE   99339982
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187445.
            Contact: Guyang Matthew Huang
            Leroy Hood
            University of Washington
            Department of Molecular Biotechnology, Box 357730, University of
            Washington, Seattle, WA 98195
            Tel: 5106280100
            Fax: 5106280108
            Email: huanggm@yahoo.com.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="tumor2"
                     /note="Organ: Prostate; Vector: pBluescript; Directional
                     cDNA library was constructed using Lambda ZP II kit
                     (Stratagene). mRNA was extracted from a frozen prostate
                     tumor  tissue (Mayo Clinics)."
BASE COUNT      172 a    130 c    156 g    134 t     44 others
ORIGIN


  Query Match            79.0%;  Score 15.8;  DB 48;  Length 636;
  Best Local Similarity  85.0%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACCG 20
           || ||||||| |||| ||||
Db     571 GCCCGTCGGTTTTCCNACCG 590


RESULT   6
C25459/c
LOCUS       C25459        212 bp    mRNA            EST       25-JUL-1997
DEFINITION  C25459 Rice shoot Oryza sativa cDNA clone S6461_1A, mRNA sequence.
ACCESSION   C25459
VERSION     C25459.1  GI:2280964
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 212)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1529555.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .212
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S6461_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       31 a     79 c     61 g     38 t      3 others
ORIGIN


  Query Match            77.0%;  Score 15.4;  DB 34;  Length 212;
  Best Local Similarity  88.9%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 GCGTCGGTATTCCGACCG 20
           ||||||||||  ||||||
Db     193 GCGTCGGTATNACGACCG 176


RESULT   7
D40400/c
LOCUS       D40400        268 bp    mRNA            EST       11-NOV-1994
DEFINITION  RICS2352A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D40400
VERSION     D40400.1  GI:569551
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 268)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .268
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       35 a     96 c     84 g     47 t      6 others
ORIGIN


  Query Match            77.0%;  Score 15.4;  DB 20;  Length 268;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 GCGTCGGTATTCCGACCG 20
           ||||||||||  ||||||
Db     190 GCGTCGGTATNACGACCG 173


RESULT   8
C25152/c
LOCUS       C25152        356 bp    mRNA            EST       25-JUL-1997
DEFINITION  C25152 Rice shoot Oryza sativa cDNA clone S5442_1A, mRNA sequence.
ACCESSION   C25152
VERSION     C25152.1  GI:2280657
KEYWORDS    EST.
```

```
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:635717.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S5442_1A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       46 a    129 c    114 g     60 t      7 others
ORIGIN


  Query Match             77.0%;  Score 15.4;  DB 34;  Length 356;
  Best Local Similarity   88.9%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 GCGTCGGTATTCCGACCG 20
          |||||||||| ||||||
Db    191 GCGTCGGTATNACGACCG 174


RESULT  9
AU057524/c
LOCUS       AU057524      570 bp    mRNA            EST       29-APR-1999
DEFINITION  AU057524 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
            clone S21530_2Z, mRNA sequence.
ACCESSION   AU057524
VERSION     AU057524.1  GI:4716408
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from mature leaf
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187981.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S21530_2Z"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT      153 a    141 c    164 g    111 t      1 others
ORIGIN


  Query Match             77.0%;  Score 15.4;  DB 49;  Length 570;
  Best Local Similarity   94.1%;  Pred. No. 1.5e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 GCGTCGGTATTCCGACC 19
          ||||||||| ||||||||
Db    103 GCGTCGGTCTTCCGACC 87


RESULT  10
FR0043767/c
LOCUS       FR0043767     595 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 173A19aD11, genomic survey
            sequence.
ACCESSION   AL131259
VERSION     AL131259.1  GI:6113205
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid clone genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Fugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 173A19"
                     /clone="173A19aD11"
BASE COUNT       78 a    148 c    139 g    115 t    115 others
ORIGIN


  Query Match             77.0%;  Score 15.4;  DB 83;  Length 595;
  Best Local Similarity   94.1%;  Pred. No. 1.6e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 GCGTCGGTATTCCGACC 19
          ||| ||||||||||||||
Db     61 GCGGCGGTATTCCGACC 45


RESULT  11
AQ648034
LOCUS       AQ648034      450 bp    DNA             GSS       22-JUN-1999
DEFINITION  RPCI93-DpnII-27F10.TV RPCI93-DpnII Trypanosoma brucei genomic clone
            RPCI93-DpnII-27F10, genomic survey sequence.
ACCESSION   AQ648034
VERSION     AQ648034.1  GI:5141220
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
```

```
REFERENCE   1  (bases 1 to 450)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
            Library for gene discovery and sequence-ready map construction
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI93-DpnII-27F10.TJ
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones and high density filters may be purchased from BACPAC
            Resources (http://bacpac.med.buffalo.edu. BAC end sequences search
            page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="RPCI93-DpnII-27F10"
                     /clone_lib="RPCI93-DpnII"
                     /note="Vector: pBACe3.6; Site_1: Bam HI; Site_2: Bam HI;
                     Constructed for The Institute for Genomic Research by
                     Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
                     Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
                     brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
                     partially digested with a combination of Eco RI and Eco RI
                     methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII
                     segment).  High molecular weight fragments were ligated in
                     pBACe3.6 vector digested with Eco RI or Bam H1,
                     respectively.  The average insert size is 141 Kb.  Total
                     coverage (both segments): > 90 X the haploid
                     non-minichromosomal genome."
BASE COUNT      114 a    119 c    122 g     95 t
ORIGIN

  Query Match             76.0%;  Score 15.2;  DB 109;  Length 450;
  Best Local Similarity   85.0%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          | ||  ||||||||||||||
Db     65 GAGCAGCGGTATTCCGACCG 84


RESULT  12
AQ808619/c
LOCUS       AQ808619      495 bp    DNA              GSS       10-AUG-1999
DEFINITION  HS_5023_A2_E02_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=599 Col=4 Row=I, genomic survey sequence.
ACCESSION   AQ808619
VERSION     AQ808619.1  GI:5727861
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3553524.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 599  row: I  column: 4
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 495.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=599 Col=4 Row=I"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      127 a    117 c    110 g    134 t      7 others
ORIGIN

  Query Match             76.0%;  Score 15.2;  DB 90;  Length 495;
  Best Local Similarity   85.0%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          || ||||||||||| |||| |
Db     36 GCTCGTCGGTATTACGACCG 17


RESULT  13
AQ448926/c
LOCUS       AQ448926      595 bp    DNA              GSS       08-APR-1999
DEFINITION  mgxb0022K04f CUGI Rice Blast BAC Library Pyricularia grisea genomic
            clone mgxb0022K04f, genomic survey sequence.
ACCESSION   AQ448926
VERSION     AQ448926.1  GI:4578063
KEYWORDS    GSS.
SOURCE      Pyricularia grisea.
  ORGANISM  Pyricularia grisea
            Eukaryota; Fungi; Ascomycota; anamorphic Ascomycota; Pyricularia.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
            Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
            Genome
  JOURNAL   Unpublished (1998)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3554972.
            Contact: Dean RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Pyricularia grisea"
                     /strain="70-15"
                     /db_xref="taxon:89476"
                     /clone="mgxb0022K04f"
                     /clone_lib="CUGI Rice Blast BAC Library"
```

/tissue_type="Protoplasts"
/lab_host="E. coli DH10B"
/note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII; Rice blast is one of the most devestating fungal diseases of rice world wide. It is a filamentous ascomycete with a haploid genome (n=7) of approximately 40 Mbp. Rice blast is an important model fungal pathogen for studying numerous aspects of the fungal-host interaction. In order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25X genome coverage. High density colony filters are available upon request."

BASE COUNT 158 a 150 c 149 g 138 t
ORIGIN

Query Match 76.0%; Score 15.2; DB 107; Length 595;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
|| |||| | ||||||||||
Db 123 GCCCGTCAGAATTCCGACCG 104

RESULT 14
AU070274
LOCUS AU070274 609 bp mRNA EST (-JUN-1999
DEFINITION AU070274 Rice mature leaf Oryza sativa cDNA clone ..1086_1A, mRNA sequence.
ACCESSION AU070274
VERSION AU070274.1 GI:5005509
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 609)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from mature leaf
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189057.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .609
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S21086_1A"
/clone_lib="Rice mature leaf"
/tissue_type="mature leaf"
BASE COUNT 102 a 183 c 206 g 116 t 2 others
ORIGIN

Query Match 76.0%; Score 15.2; DB 51; Length 609;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
|||| ||||| ||||||||
Db 445 GCGCACCGGTACTCCGACCG 464

RESULT 15
AU057624
LOCUS AU057624 665 bp mRNA EST 29-APR-1999
DEFINITION AU057624 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA clone S21626_2Z, mRNA sequence.
ACCESSION AU057624
VERSION AU057624.1 GI:4716508
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 665)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from mature leaf
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188081.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .665
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S21626_2Z"
/clone_lib="Oryza sativa mature leaf Nipponbare"
/tissue_type="mature leaf"
BASE COUNT 182 a 137 c 173 g 170 t 3 others
ORIGIN

Query Match 76.0%; Score 15.2; DB 49; Length 665;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
|||| ||||| ||||||||
Db 153 GCGCACCGGTACTCCGACCG 172

Search completed: June 22, 2000, 12:08:16
Job time: 10410 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:32 ; Search time 341.15 Seconds
(without alignments)
7.620 Million cell updates/sec

Title: US-09-362-485-12
Perfect score: 20
Sequence: 1 GCGCGTCGGTATTCCGACCG 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.4	72.0	2310	2	US-08-525-654A-2	Sequence 2, Appli
c	2	14.4	72.0	2313	2	US-08-525-654A-4	Sequence 4, Appli
c	3	14.4	72.0	3394	2	US-08-525-654A-136	Sequence 136, App
	4	14.2	71.0	2471	2	US-08-920-812-14	Sequence 14, Appl
	5	14.2	71.0	2471	2	US-08-920-827-14	Sequence 14, Appl
	6	14.2	71.0	2471	2	US-08-921-177-14	Sequence 14, Appl
	7	14.2	71.0	2471	2	US-08-362-577C-14	Sequence 14, Appl
	8	14.2	71.0	2471	3	US-08-920-828-14	Sequence 14, Appl
	9	14.2	71.0	3836	3	US-08-216-260-1	Sequence 1, Appli
	10	14.2	71.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
	11	14.2	71.0	28958	1	US-08-456-837-6	Sequence 6, Appli
	12	14.2	71.0	28958	1	US-08-457-342-6	Sequence 6, Appli
	13	14.2	71.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
	14	14.2	71.0	28958	2	US-08-458-076A-6	Sequence 6, Appli
	15	14.2	71.0	28958	2	US-08-764-233A-4	Sequence 4, Appli
	16	14.2	71.0	28958	2	US-08-457-335A-6	Sequence 6, Appli
	17	14.2	71.0	28958	2	US-08-729-214-6	Sequence 6, Appli
	18	14.2	71.0	49377	2	US-08-764-233A-1	Sequence 1, Appli
c	19	13.6	68.0	311	5	US-08-937-580-20	Sequence 20, Appl
	20	13.6	68.0	1005	2	US-08-482-385A-2	Sequence 2, Appli
	21	13.6	68.0	2728	2	US-08-482-385A-5	Sequence 5, Appli
c	22	13.4	67.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	23	13.4	67.0	28958	1	US-08-456-837-6	Sequence 6, Appli
c	24	13.4	67.0	28958	1	US-08-457-342-6	Sequence 6, Appli
c	25	13.4	67.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	26	13.4	67.0	28958	2	US-08-458-076A-6	Sequence 6, Appli
c	27	13.4	67.0	28958	2	US-08-764-233A-4	Sequence 4, Appli
c	28	13.4	67.0	28958	2	US-08-457-335A-6	Sequence 6, Appli
c	29	13.4	67.0	28958	2	US-08-729-214-6	Sequence 6, Appli
c	30	13.4	67.0	49377	2	US-08-764-233A-1	Sequence 1, Appli
c	31	13.2	66.0	311	5	US-08-937-580-19	Sequence 19, Appl
	32	13.2	66.0	1362	1	US-07-956-700B-5	Sequence 5, Appli
	33	13.2	66.0	1362	2	US-08-476-537-5	Sequence 5, Appli
	34	13.2	66.0	1362	2	US-08-485-607-5	Sequence 5, Appli
	35	13.2	66.0	1362	2	US-08-611-107-7	Sequence 7, Appli
	36	13.2	66.0	1362	4	US-08-422-560A-7	Sequence 7, Appli
	37	13.2	66.0	1362	4	US-08-475-879-5	Sequence 5, Appli
c	38	13.2	66.0	1448	1	US-08-035-634-1	Sequence 1, Appli
	39	13.2	66.0	2444	5	US-08-920-610-1	Sequence 1, Appli
c	40	12.8	64.0	324	4	US-08-690-011A-22	Sequence 22, Appl
	41	12.8	64.0	561	3	US-08-655-821-1	Sequence 1, Appli
c	42	12.8	64.0	660	3	US-08-401-068-5	Sequence 5, Appli
c	43	12.8	64.0	660	3	US-08-846-338-5	Sequence 5, Appli
c	44	12.8	64.0	15239	2	US-08-390-878-17	Sequence 17, Appl
	45	12.6	63.0	560	1	US-07-947-120-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-08-525-654A-2/c
; Sequence 2, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
;    TITLE OF INVENTION:  TRANSGLUTAMINASE ORIGINATED FROM
;    TITLE OF INVENTION:  CRASSOSTREA GIGAS
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,654A
;      FILING DATE:  28-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6/8283
;      FILING DATE:  28-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7/3876
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-760-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2310 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Crassostrea gigas
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2310
US-08-525-654A-2


  Query Match             72.0%;  Score 14.4;  DB 2;  Length 2310;
  Best Local Similarity   93.8%;  Pred. No. 24;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 CGCGTCGGTATTCCGA 17
           || |||||||||||||
Db    1091 CGGGTCGGTATTCCGA 1076


RESULT   2
US-08-525-654A-4/c
; Sequence 4, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
;    TITLE OF INVENTION:  TRANSGLUTAMINASE ORIGINATED FROM
;    TITLE OF INVENTION:  CRASSOSTREA GIGAS
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,654A
;      FILING DATE:  28-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6/8283
;      FILING DATE:  28-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7/3876
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-760-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2313 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Crassostrea gigas
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2054
US-08-525-654A-4


  Query Match             72.0%;  Score 14.4;  DB 2;  Length 2313;
  Best Local Similarity   93.8%;  Pred. No. 24;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 CGCGTCGGTATTCCGA 17
           || |||||||||||||
Db    1094 CGGGTCGGTATTCCGA 1079


RESULT   3
US-08-525-654A-136/c
; Sequence 136, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
;    TITLE OF INVENTION:  TRANSGLUTAMINASE ORIGINATED FROM
;    TITLE OF INVENTION:  CRASSOSTREA GIGAS
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,654A
;      FILING DATE:  28-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6/8283
;      FILING DATE:  28-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7/3876
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-760-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  136:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3394 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Crassostrea gigas
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  305..2617
US-08-525-654A-136


  Query Match             72.0%;  Score 14.4;  DB 2;  Length 3394;
  Best Local Similarity   93.8%;  Pred. No. 24;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      2 CGCGTCGGTATTCCGA 17
          || ||||||||||||
Db   1398 CGGGTCGGTATTCCGA 1383


RESULT   4
US-08-920-812-14
; Sequence 14, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-812-14


  Query Match            71.0%;  Score 14.2;  DB 2;  Length 2471;
  Best Local Similarity  84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACC 19
          |||| ||| |||| |||||
Db   1632 GCGCTTCGATATTGCGACC 1650


RESULT   5
US-08-920-827-14
; Sequence 14, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-827-14


  Query Match            71.0%;  Score 14.2;  DB 2;  Length 2471;
  Best Local Similarity  84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACC 19
          |||| ||| |||| |||||
Db   1632 GCGCTTCGATATTGCGACC 1650


RESULT   6
US-08-921-177-14
; Sequence 14, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-921-177-14


  Query Match             71.0%;  Score 14.2;  DB 2;  Length 2471;
  Best Local Similarity   84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||| |||| |||||
Db    1632 GCGCTTCGATATTGCGACC 1650


RESULT   7
US-08-362-577C-14
; Sequence 14, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-362-577C-14


  Query Match             71.0%;  Score 14.2;  DB 2;  Length 2471;
  Best Local Similarity   84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||| |||| |||||
Db    1632 GCGCTTCGATATTGCGACC 1650


RESULT   8
US-08-920-828-14
; Sequence 14, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-828-14
```

```
  Query Match           71.0%;  Score 14.2;  DB 3;  Length 2471;
  Best Local Similarity  84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||| |||| |||||
Db    1632 GCGCTTCGATATTGCGACC 1650


RESULT   9
US-08-216-260-1
; Sequence 1, Application US/08216260
; Patent No. 5837261
;  GENERAL INFORMATION:
;    APPLICANT:  Inglis, Stephen C.
;    APPLICANT:  Boursnell, Michael E.G.
;    APPLICANT:  Minson, Anthony C.
;    TITLE OF INVENTION:  VIRAL VACCINES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      ADDRESSEE:  Attn: Walter H. Dreger
;      STREET:  4 Embarcadero Center, Suite 340(
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/216,260
;      FILING DATE:  21-MAR-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9305710.7
;      FILING DATE:  19-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9324964.7
;      FILING DATE:  06-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/030,073
;      FILING DATE:  19-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/168,643
;      FILING DATE:  16-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Walter H.
;      REGISTRATION NUMBER:  24,190
;      REFERENCE/DOCKET NUMBER:  A-59886/WHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3836 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
US-08-216-260-1


  Query Match           71.0%;  Score 14.2;  DB 3;  Length 3836;
  Best Local Similarity  84.2%;  Pred. No. 31;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 CGCGTCGGTATTCCGACCG 20
           || ||||| | ||||||||
Db    1605 CGGGTCGGAAATCCGACCG 1623


RESULT  10
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,261B
;      FILING DATE:  08-JUN-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6


  Query Match           71.0%;  Score 14.2;  DB 1;  Length 28958;
  Best Local Similarity  84.2%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504


RESULT  11
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
```

```
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6


  Query Match             71.0%;  Score 14.2;  DB 1;  Length 28958;
  Best Local Similarity   84.2%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504


RESULT  12
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6


  Query Match             71.0%;  Score 14.2;  DB 1;  Length 28958;
  Best Local Similarity   84.2%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504


RESULT  13
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
```

```
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6


  Query Match             71.0%;  Score 14.2;  DB 1;  Length 28958;
  Best Local Similarity   84.2%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504


RESULT  14
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6


  Query Match             71.0%;  Score 14.2;  DB 2;  Length 28958;
  Best Local Similarity   84.2%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504


RESULT  15
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
```

; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1
US-08-764-233A-4

Query Match 71.0%; Score 14.2; DB 2; Length 28958;
Best Local Similarity 84.2%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 GCGCGTCGGTATTCCGACC 19
           |||| ||  ||||||||||
Db   12486 GCGCCTCCATATTCCGACC 12504
```

Search completed: June 22, 2000, 15:21:36
Job time: 18774 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:38 ; Search time 458.59 Seconds
(without alignments)
10.911 Million cell updates/sec

Title: US-09-362-485-12
Perfect score: 20
Sequence: 1 GCGCGTCGGTATTCCGACCG 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	100.0	20	1	V49513	Mycobacterium sp.
	2	20	100.0	20	1	V49611	AlaDH derived olig
	3	20	100.0	1245	1	V49626	Mycobacterium tube
	4	20	100.0	1260	1	V49510	Mycobacterium sp.
	5	20	100.0	1260	1	V49625	Mycobacterium tube
	6	16	80.0	18	1	V49512	Mycobacterium sp.
	7	16	80.0	18	1	V49610	AlaDH derived olig
	8	15.2	76.0	804	1	Q98295	Recombinant abrin
	9	15.2	76.0	804	1	Q98296	Recombinant abrin
	10	15.2	76.0	807	1	T10398	Recombinant abrin
	11	15.2	76.0	1439	1	V44424	Mycobacterium tube
	12	15.2	76.0	1439	1	V64533	M. tuberculosis im
c	13	15.2	76.0	4800	1	X05350	Phenol hydroxylase
c	14	15.2	76.0	4800	1	X01590	Pseudomonas putida
c	15	14.4	72.0	3394	1	Q99223	Japanese oyster tr
	16	14.4	72.0	29555	1	V74517	Staphylococcus aur
c	17	14.2	71.0	93	1	Q72987	Fragile X syndrome
c	18	14.2	71.0	370	1	V20274	Probe (175) for mi
	19	14.2	71.0	1028	1	Q27091	XTY26 probe. DNA s
	20	14.2	71.0	2325	1	T95400	Flavobacterium odo
	21	14.2	71.0	2438	1	T47765	Comamonas acidovor
	22	14.2	71.0	2471	1	Q55146	Pseudomonas aerugi
	23	14.2	71.0	3836	1	Q73652	HSV-2 strain 25766
	24	14.2	71.0	3836	1	V64102	HSV-2 strain 25766
	25	14.2	71.0	5222	1	Q31891	Fragile X region c
c	26	14.2	71.0	20387	1	V62159	HSV-2 strain SB5 C
	27	14.2	71.0	25187	1	X20534	Polynucleotide seq
c	28	14.2	71.0	26338	1	V62134	HSV-2 strain SB5 C
	29	14.2	71.0	28598	1	T06769	Sorangium cellulos
	30	14.2	71.0	28958	1	T89956	Sorangium cellulos
	31	14.2	71.0	49377	1	V05287	The soraphen biosy
c	32	14.2	71.0	117213	1	V62176	HSV-2 strain SB5 C
c	33	13.8	69.0	1038	1	V84117	Pseudomonas aerugi
c	34	13.8	69.0	1518	1	T37313	Aromatic acyl tran
	35	13.8	69.0	1935	1	X13361	Enterococcus faeca
c	36	13.8	69.0	5420	1	X20601	Polynucleotide seq
c	37	13.6	68.0	633	1	T63512	Urease accessory m
	38	13.6	68.0	1005	1	Q83794	S. avermitilis BCK
c	39	13.6	68.0	1050	1	X34044	Mycobacterium spec
c	40	13.6	68.0	1269	1	X34045	Mycobacterium spec
	41	13.6	68.0	1564	1	Q38955	Lipase gene #1. DN
c	42	13.6	68.0	1705	1	T17868	Murine interleukin
c	43	13.6	68.0	1714	1	T32613	Murine Etl-2 gene.
	44	13.6	68.0	2232	1	Q89835	Human death associ
	45	13.6	68.0	2232	1	V60288	DNA sequence encod

ALIGNMENTS

RESULT 1
V49513
ID V49513 standard; DNA; 20 BP.
AC V49513;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F1+.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 20 BP; 2 A; 7 C; 7 G; 4 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
||||||||||||||||||||
Db 1 GCGCGTCGGTATTCCGACCG 20

RESULT 2
V49611
ID V49611 standard; DNA; 20 BP.
AC V49611;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F1+.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 20 BP; 2 A; 7 C; 7 G; 4 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
||||||||||||||||||||
Db 1 GCGCGTCGGTATTCCGACCG 20

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 20; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
||||||||||||||||||||
Db 63 GCGCGTCGGTATTCCGACCG 82

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 2 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 20; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGTCGGTATTCCGACCG 20
||||||||||||||||||||
Db 78 GCGCGTCGGTATTCCGACCG 97

RESULT 5
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match            100.0%;  Score 20;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 0.047;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     78 GCGCGTCGGTATTCCGACCG 97


RESULT   6
V49512
ID   V49512 standard; DNA; 18 BP.
AC   V49512;
DT   20-OCT-1998  (first entry)
DE   Mycobacterium sp. AlaDH oligonucleotide AlaDH-F1.
KW   Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW   swimmers disease; vaccine; epidemic; infection; identification; ss.
OS   Synthetic.
OS   Mycobacterium sp.
PN   WO9832862-A2.
PD   30-JUL-1998.
PF   29-JAN-1998; E00484.
PR   29-JAN-1997; EP-101339.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-427958/36.
PT   Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT   - used for diagnosis of tuberculosis and other mycobacterial
PT   diseases, also for treatment and prevention, for drug screening and
PT   for bio-transformation
PS   Disclosure; Page 10; 57pp; German.
CC   V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC   dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC   to diagnose tuberculosis and other mycobacterial infections (including
CC   'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC   animals. The protein can also be used for contro. of epidemics and for
CC   vaccination, to screen for agents with anti-mycobacterial activity, and
CC   in bio-transformations that are specific for L-alanine. Also mycobacteria
CC   can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC   that is secreted early during infection.
SQ   Sequence   18 BP;    2 A;    5 C;    6 G;    5 T;


  Query Match            80.0%;  Score 16;  DB 1;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 5.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCG 16
          ||||||||||||||||
Db      3 GCGCGTCGGTATTCCG 18


RESULT   7
V49610
ID   V49610 standard; DNA; 18 BP.
AC   V49610;
DT   20-NOV-1998  (first entry)
DE   AlaDH derived oligonucleotide AlaDH-F1.
KW   ss; Alanine dehydrogenase; tuberculosis; TB; ycobacteria; L-alanine.
OS   Mycobacterium sp.
PN   WO9836089-A2.
PD   20-AUG-1998.
PF   29-JAN-1998; E00483.
PR   29-JAN-1997; EP-101338.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-457123/39.
PT   Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT   - useful for, e.g. for diagnosis, differentiation of strains,
PT   monitoring vaccination and identification of mycobacterial
PT   inhibitors
PS   Claim 9; Fig 2.5; 55pp; German.
CC   THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC   are used for diagnosing tuberculosis (TB) and other mycobacterial
CC   infections in humans or animals.  Kits are used for direct diagnosis of
CC   TB on clinical samples (  g. body fluids) and can differentiate between
CC   pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC   the M.tuberculosis (M. t.) complex.  The kit may also be used to
CC   identify substances that inhibit mycobacteria, for combatting epidemics
CC   and for vaccination follow-up.  The oligonucleotides are used similarly
CC   in diagnostic hybridisation tests, also for culture confirmation of
CC   isolated strains and for chromosome fingerprinting to detect/
CC   differentiate between mycobacteria, and for L-alanine-specific
CC   biotransformation reactions.  AlaDH is an early antigen, present
CC   extracellularly after only a few days of growth, making it an ideal
CC   drug target.
SQ   Sequence   18 BP;    2 A;    5 C;    6 G;    5 T;


  Query Match            80.0%;  Score 16;  DB 1;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 5.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCG 16
          ||||||||||||||||
Db      3 GCGCGTCGGTATTCCG 18


RESULT   8
Q98295
ID   Q98295 standard; DNA; 804 BP.
AC   Q98295;
DT   20-MAR-1996  (first entry)
DE   Recombinant abrin A variant with Ser202.
KW   Recombinant abrin A variant; E.coli; cross-link; antibody; Fab';
KW   antibody-toxicity complex; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             27. .785
FT                   /*tag= a
FT                   /product= recombinant abrin A
FT   mat_peptide     30. .782
FT                   /*tag= b
FT   misc_difference 435. .437
FT                   /*tag= c
FT                   /transl_except= seq: CAC,  a.a.:Glu
FT   misc_difference 438. .440
FT                   /*tag= d
FT                   /transl_except= seq: GGT,  a.a.:Glu
FT   misc_difference 441. .443
FT                   /*tag= e
FT                   /transl_except= seq: ATC,  a.a.:Lys
FT   misc_difference 444. .446
FT                   /*tag= f
FT                   /transl_except= seq: TCA,  a.a.:His
FT   misc_difference 447. .449
FT                   /*tag= g
FT                   /transl_except= seq: TTT,  a.a.:Gly
FT   misc_difference 450. .452
FT                   /*tag= h
FT                   /transl_except= seq: TTT,  a.a.:Ile
FT   misc_difference 453. .455
FT                   /*tag= i
FT                   /transl_except= seq: CGA,  a.a.:Ser
FT   misc_difference 456. .458
```

```
FT                   /*tag= j
FT                   /transl_except= seq: AGT,  a.a.:Phe
FT   misc_difference 459. .461
FT                   /*tag= k
FT                   /transl_except= seq: GGT,  a.a.:Phe
FT   misc_difference 462. .464
FT                   /*tag= l
FT                   /transl_except= seq: GGC,  a.a.:Arg
FT   misc_difference 465. .467
FT                   /*tag= m
FT                   /transl_except= seq: AAC,  a.a.:Ser
FT   misc_difference 468. .470
FT                   /*tag= n
FT                   /transl_except= seq: GAC,  a.a.:Gly
FT   misc_difference 471. .473
FT                   /*tag= o
FT                   /transl_except= seq: AAC,  a.a.:Gly
FT   misc_difference 474. .476
FT                   /*tag= p
FT                   /transl_except= seq: GAG,  a.a.:Asn
FT   misc_difference 477. .479
FT                   /*tag= q
FT                   /transl_except= seq: GAA,  a.a.:Asp
FT   misc_difference 480. .482
FT                   /*tag= r
FT                   /transl_except= seq: AAA,  a.a.:Asn
PN   J07188290-A.
PD   25-JUL-1995.
PF   27-DEC-1993; 331357.
PR   27-DEC-1993; JP-331357.
PA   (TORA ) TORAY IND INC.
DR   WPI; 95-290424/38.
DR   P-PSDB; R81976.
PT   Recombinant abrin A chain variant - useful with crosslinked moieties
PT   eg. antibodies, to form a starting material for pharmaceuticals.
PS   Example 1; Page 7-8; 13pp; Japanese.
CC   The nucleotide sequence of the recombinant abrin A variant gene encoding
CC   a 252 amino acid protein contg. a Ser residue at pos. 202.  The
CC   recombinant DNA is useful for producing large amounts of highly active
CC   abrin A at high purity, pref. E.coli BL21.  The can be cross-linked to an
CC   antibody, pref. an Fab' fragment or an antigen molecule.  When
CC   cross-linked to an antibody molecule, the protein used as the starting
CC   material for antibody-toxicity complexes, etc. in pharmaceutical
CC   compositions.
SQ   Sequence   804 BP;    211 A;    208 C;    175 G;    210 T;

  Query Match            76.0%;  Score 15.2;  DB 1;  Length 804;
  Best Local Similarity  85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||  |||||| |||||
Db    554 GCGCGTTAGTATTCAGACCG 573


RESULT   9
Q98296
ID   Q98296 standard; DNA; 804 BP.
AC   Q98296;
DT   20-MAR-1996  (first entry)
DE   Recombinant abrin A variant with Ala202.
KW   Recombinant abrin A variant; E.coli; cross-link; antibody; Fab';
KW   antibody-toxicity complex; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             27. .785
FT                   /*tag= a
FT                   /product= recombinant abrin A
FT   mat_peptide     30. .782
FT                   /*tag= b
FT   misc_difference 435. .437
FT                   /*tag= c
FT                   /transl_except= seq: CAC,  a.a.:Glu
FT   misc_difference 438. .440
FT                   /*tag= d
FT                   /transl_except= seq: GGT,  a.a.:Glu
FT   misc_difference 441. .443
FT                   /*tag= e
FT                   /transl_except= seq: ATC,  a.a.:Lys
FT   misc_difference 444. .446
FT                   /*tag= f
FT                   /transl_except= seq: TCA,  a.a.:His
FT   misc_difference 447. .449
FT                   /*tag= g
FT                   /transl_except= seq: TTT,  a.a.:Gly
FT   misc_difference 450. .452
FT                   /*tag= h
FT                   /transl_except= seq: TTT,  a.a.:Ile
FT   misc_difference 453. .455
FT                   /*tag= i
FT                   /transl_except= seq: CGA,  a.a.:Ser
FT   misc_difference 456. .458
FT                   /*tag= j
FT                   /transl_except= seq: AGT,  a.a.:Phe
FT   misc_difference 459. .461
FT                   /*tag= k
FT                   /transl_except= seq: GGT,  a.a.:Phe
FT   misc_difference 462. .464
FT                   /*tag= l
FT                   /transl_except= seq: GGC,  a.a.:Arg
FT   misc_difference 465. .467
FT                   /*tag= m
FT                   /transl_except= seq: AAC,  a.a.:Ser
FT   misc_difference 468. .470
FT                   /*tag= n
FT                   /transl_except= seq: GAC,  a.a.:Gly
FT   misc_difference 471. .473
FT                   /*tag= o
FT                   /transl_except= seq: AAC,  a.a.:Gly
FT   misc_difference 474. .476
FT                   /*tag= p
FT                   /transl_except= seq: GAG,  a.a.:Asn
FT   misc_difference 477. .479
FT                   /*tag= q
FT                   /transl_except= seq: GAA,  a.a.:Asp
FT   misc_difference 480. .482
FT                   /*tag= r
FT                   /transl_except= seq: AAA,  a.a.:Asn
PN   J07188290-A.
PD   25-JUL-1995.
PF   27-DEC-1993; 331357.
PR   27-DEC-1993; JP-331357.
PA   (TORA ) TORAY IND INC.
DR   WPI; 95-290424/38.
DR   P-PSDB; R81977.
PT   Recombinant abrin A chain variant - useful with crosslinked moieties
PT   eg. antibodies, to form a starting material for pharmaceuticals.
PS   Example 1; Page 8-9; 13pp; Japanese.
CC   The nucleotide sequence of the recombinant abrin A variant gene encoding
CC   a 252 amino acid protein contg. an Ala residue at pos. 202.  The
CC   recombinant DNA is useful for producing large amounts of highly active
CC   abrin A at high purity, pref. E.coli BL21.  The can be cross-linked to an
CC   antibody, pref. an Fab' fragment or an antigen molecule.  When
CC   cross-linked to an antibody molecule, the protein used as the starting
CC   material for antibody-toxicity complexes, etc. in pharmaceutical
CC   compositions.
SQ   Sequence   804 BP;    210 A;    210 C;    176 G;    208 T;

  Query Match            76.0%;  Score 15.2;  DB 1;  Length 804;
  Best Local Similarity  85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||  |||||| |||||
```

```
Db    554 GCGCGTTAGTATTCAGACCG 573


RESULT  10
T10398
ID   T10398 standard; DNA; 807 BP.
AC   T10398;
DT   20-MAR-1996  (first entry)
DE   Recombinant abrin A variant with Ser203.
KW   Recombinant abrin A variant; E.coli; cross-link; antibody; Fab';
KW   antibody-toxicity complex; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             27. .788
FT                   /*tag= a
FT                   /product= recombinant abrin A
FT   mat_peptide     30. .785
FT                   /*tag= b
FT   misc_difference 438. .440
FT                   /*tag= c
FT                   /transl_except= seq: CAC,  a.a.:Glu
FT   misc_difference 441. .443
FT                   /*tag= d
FT                   /transl_except= seq: GGT,  a.a.:Glu
FT   misc_difference 444. .446
FT                   /*tag= e
FT                   /transl_except= seq: ATC,  a.a.:Lys
FT   misc_difference 447. .449
FT                   /*tag= f
FT                   /transl_except= seq: TCA,  a.a.:His
FT   misc_difference 450. .452
FT                   /*tag= g
FT                   /transl_except= seq: TTT,  a.a.:Gly
FT   misc_difference 453. .455
FT                   /*tag= h
FT                   /transl_except= seq: TTT,  a.a.:Ile
FT   misc_difference 456. .458
FT                   /*tag= i
FT                   /transl_except= seq: CGA,  a.a.:Ser
FT   misc_difference 459. .461
FT                   /*tag= j
FT                   /transl_except= seq: AGT,  a.a.:Phe
FT   misc_difference 462. .464
FT                   /*tag= k
FT                   /transl_except= seq: GGT,  a.a.:Phe
FT   misc_difference 465. .467
FT                   /*tag= l
FT                   /transl_except= seq: GGC,  a.a.:Arg
FT   misc_difference 468. .470
FT                   /*tag= m
FT                   /transl_except= seq: AAC,  a.a.:Ser
FT   misc_difference 471. .473
FT                   /*tag= n
FT                   /transl_except= seq: GAC,  a.a.:Gly
FT   misc_difference 474. .476
FT                   /*tag= o
FT                   /transl_except= seq: AAC,  a.a.:Gly
FT   misc_difference 477. .479
FT                   /*tag= p
FT                   /transl_except= seq: GAG,  a.a.:Asn
FT   misc_difference 480. .482
FT                   /*tag= q
FT                   /transl_except= seq: GAA,  a.a.:Asp
FT   misc_difference 483. .485
FT                   /*tag= r
FT                   /transl_except= seq: AAA,  a.a.:Asn
PN   J07188290-A.
PD   25-JUL-1995.
PF   27-DEC-1993; 331357.
PR   27-DEC-1993; JP-331357.
PA   (TORA ) TORAY IND INC.
DR   WPI; 95-290424/38.
DR   P-PSDB; R81978.
PT   Recombinant abrin A chain variant - useful with crosslinked moieties
PT   eg. antibodies, to form a starting material for pharmaceuticals.
PS   Example 1; Page 9-10; 13pp; Japanese.
CC   The nucleotide sequence of the recombinant abrin A variant gene encoding
CC   a 252 amino acid protein contg. a Ser residue at pos. 203.  The
CC   recombinant DNA is useful for producing large amounts of highly active
CC   abrin A at high purity, pref. E.coli BL21.  The can be cross-linked to an
CC   antibody, pref. an Fab' fragment or an antigen molecule.  When
CC   cross-linked to an antibody molecule, the protein used as the starting
CC   material for antibody-toxicity complexes, etc. in pharmaceutical
CC   compositions.
SQ   Sequence   807 BP;    210 A;    209 C;    177 G;    211 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 807;
  Best Local Similarity  85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||  |||||| ||||| 
Db    557 GCGCGTTAGTATTCAGACCG 576


RESULT  11
V44424
ID   V44424 standard; DNA; 1439 BP.
AC   V44424;
DT   09-NOV-1998  (first entry)
DE   Mycobacterium tuberculosis antigen XP17 DNA.
KW   Tuberculosis; infection; diagnosis; antigen; XP17; ss.
OS   Mycobacterium tuberculosis strain Erdman.
PN   WO9816645-A1.
PD   23-APR-1998.
PF   07-OCT-1997; U18214.
PR   13-MAR-1997; US-818111.
PR   11-OCT-1996; US-729622.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-251292/22.
PT   New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and diagnosis of tuberculosis
PS   Claim 4; Page 179-180; 250pp; English.
CC   This DNA sequence codes for an antigenic portion of Mycobacterium
CC   tuberculosis antigen XP17.  It was isolated from a M. tuberculosis
CC   strain Erdman genomic DNA expression library using sera from
CC   patients having extrapulmonary tuberculosis.  The clone bears no
CC   similarity to known sequences.  The invention relates to methods
CC   and compositions for diagnosing tuberculosis.  It provides
CC   polypeptides (see W64291-W64379) comprising an antigenic portion of
CC   a soluble M. tuberculosis antigen, or an immunogenic portion of a
CC   M. tuberculosis antigen, as well as DNA sequences encoding such
CC   polypeptides, recombinant expression vectors and transformed or
CC   transfected host cells.  Also claimed are methods and diagnostic
CC   kits for detecting M. tuberculosis infection in a patient using
CC   these polypeptides, antibodies or oligonucleotide probes and
CC   primers.
SQ   Sequence   1439 BP;    265 A;    444 C;    529 G;    201 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 1439;
  Best Local Similarity  85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          |||| ||||| ||||| |||
Db    548 GCGCTTCGGTGTTCCGCCCG 567


RESULT  12
V64533
ID   V64533 standard; DNA; 1439 BP.
```

```
AC   V64533;
DT   27-JAN-1999  (first entry)
DE   M. tuberculosis immunogenic polypeptide XP17 5'-end DNA.
KW   Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW   vaccine; pharmaceutical; infection; diagnosis; ss.
OS   Mycobacterium tuberculosis.
PN   WO9816646-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18293.
PR   13-MAR-1997; US-818112.
PR   11-OCT-1996; US-730510.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-261042/23.
PT   Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and for diagnosis, treatment and prevention of tuberculosis
PS   Claim 4; Page 159; 230pp; English.
CC   This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC   tuberculosis (MT) antigen which can be used in a method for inducing
CC   protective immunity against tuberculosis (TB). This sequence can be
CC   formulated into vaccines and/or pharmaceutical compositions for
CC   immunising against M. tuberculosis infection or may be used for the
CC   diagnosis of tuberculosis.
SQ   Sequence   1439 BP;    265 A;    444 C;    529 G;    201 T;


  Query Match             76.0%;  Score 15.2;  DB 1;  Length 1439;
  Best Local Similarity   85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACCG 20
           |||| ||||| ||||| |||
Db     548 GCGCTTCGGTGTTCCGCCCG 567


RESULT  13
X05350/c
ID   X05350 standard; DNA; 4800 BP.
AC   X05350;
DT   16-APR-1999  (first entry)
DE   Phenol hydroxylase genomic DNA sequence.
KW   Phenol hydroxylase; recombinant; phenol-assimilating microbe;
KW   chlorinated; ethylene; decomposition; aerobical; ds.
OS   Pseudomonas putida.
FH   Key             Location/Qualifiers
FT   CDS             127. .345
FT                   /*tag= a
FT                   /gene= "pheZ"
FT   CDS             434. .1429
FT                   /*tag= b
FT                   /gene= "pheA"
FT   CDS             1440. .1712
FT                   /*tag= c
FT                   /gene= "pheB"
FT   CDS             1754. .3268
FT                   /*tag= d
FT                   /gene= "pheC"
FT   CDS             3301. .3660
FT                   /*tag= e
FT                   /gene= "pheD"
FT   CDS             3689. .4756
FT                   /*tag= f
FT                   /gene= "pheE"
PN   J11018773-A.
PD   26-JAN-1999.
PF   30-JUN-1997; 174996.
PR   30-JUN-1997; JP-174996.
PA   (KURK ) KURITA WATER IND LTD.
DR   WPI; 99-161080/14.
PT   Gene recombinant decomposing chlorinated ethylene  - has a promoter
PT   inserted upstream of a phenol hydroxylase gene
PS   Disclosure; Page 8-10; 12pp; Japanese.
CC   The present sequence represents a phenol hydroxylase genomic DNA sequence
CC   isolated from Pseudomonas putida strain KWI-9. The invention relates to a
CC   recombinant gene carrying a recombinant DNA in which a promoter is
CC   inserted upstream of a phenol hydroxylase gene derived from a phenol-
CC   assimilating microbe having chlorinated ethylene-decomposing activity.
CC   The said phenol hydroxylase gene is a gene gp. consisting of chlorinated
CC   ethylene-decomposition-promoting gene and chlorinated ethylene-
CC   decomposing gene gp. and carries said recombinant DNA on a chromosome.
CC   The invention also provides a method for the decomposition of chlorinated
CC   ethylene using the above gene recombinant. The gene recombinant gives a
CC   high decomposition of chlorinated ethylene per microbe body. The method
CC   decomposes chlorinated ethylene efficiently under aerobical condition.
SQ   Sequence   4800 BP;    897 A;    1579 C;    1534 G;    790 T;


  Query Match             76.0%;  Score 15.2;  DB 1;  Length 4800;
  Best Local Similarity   85.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GCGCGTCGGTATTCCGACCG 20
           |||| ||||| ||||| |||
Db     48 GCGCGTCGGGATTCTGGCCG 29


RESULT  14
X01590/c
ID   X01590 standard; DNA; 4800 BP.
AC   X01590;
DT   05-MAY-1999  (first entry)
DE   Pseudomonas putida phe operon.
KW   Chlorinated ethylene decomposition; pheZ; pheA; pheB; pheC; pheD; pheE;
KW   chlorinated ethylene-decomposing gene; ds.
OS   Pseudomonas putida.
FH   Key             Location/Qualifiers
FT   CDS             127. .345
FT                   /*tag= a
FT                   /product= pheZ
FT   CDS             434. .1429
FT                   /*tag= b
FT                   /product= pheA
FT   CDS             1440. .1712
FT                   /*tag= c
FT                   /product= pheB
FT   CDS             1754. .3268
FT                   /*tag= d
FT                   /product= pheC
FT   CDS             3301. .3660
FT                   /*tag= e
FT                   /product= pheD
FT   CDS             3689. .4756
FT                   /*tag= f
FT                   /product= pheE
PN   J11018774-A.
PD   26-JAN-1999.
PF   30-JUN-1997; 174997.
PR   30-JUN-1997; JP-174997.
PA   (KURK ) KURITA WATER IND LTD.
DR   WPI; 99-161081/14.
PT   Gene recombinant decomposing chlorinated ethylene - has DNA fragment
PT   which stops transcription of a gene downstream of chlorinated
PT   ethylene decomposing gene
PS   Disclosure; Page 8-10; 12pp; Japanese.
CC   This sequence represents the Pseudomonas putida strain KWI-9 pheZ, pheA,
CC   pheB, pheC, pheD, pheE coding sequence. The invention relates to a gene
CC   recombinant capable of decomposing chlorinated ethylene characterised by
CC   that a recombinant DNA in which a DNA fragment having a function of
CC   stopping transcription of a gene is inserted downstream of a chlorinated
CC   ethylene-decomposing gene derived from a microbe having chlorinated
CC   ethylene-decomposing activity is carried on a chromosome. The gene
CC   recombinant gives a high decomposition of chlorinated ethylene per
CC   microbe body.  It also decomposes chlorinated ethylene efficiently under
CC   aerobic condition.
```

SQ Sequence 4800 BP; 897 A; 1579 C; 1534 G; 790 T;

Query Match 76.0%; Score 15.2; DB 1; Length 4800;
Best Local Similarity 85.0%; Pred. No. 17;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 GCGCGTCGGTATTCCGACCG 20
           |||||||||  |||| | |||
Db      48 GCGCGTCGGGATTCTGGCCG 29
```

RESULT 15
Q99223/c
ID Q99223 standard; cDNA to mRNA; 3394 BP.
AC Q99223;
DT 06-MAR-1996 (first entry)
DE Japanese oyster transglutaminase cDNA.
KW Japanese oyster; transglutaminase; gelling agent;
KW yoghurt; jelly; cheese; fish-paste; calcium ion activation; ds.
OS Crassostrea gigas.
FH Key Location/Qualifiers
FT cds 305. .2620
FT /*tag= a
PN WO9520662-A1.
PD 03-AUG-1995.
PF 30-JAN-1995; J00117.
PR 28-JAN-1994; JP-008283.
PR 13-JAN-1995; JP-003878.
PA (AJIN) AJINOMOTO CO INC.
PI Kumazawa Y, Motoki M, Sano K, Seguro K, Yasueda H;
DR WPI; 95-275447/36.
DR P-PSDB; R80495.
PT Transglutaminase derived from the Japanese oyster - is activated by
PT calcium ions and is a gelling agent for foodstuffs
PS Claim 5; Pages 97-105; 127pp; Japanese.
CC Q99223 encodes R80495 Japanese oyster transglutaminase (TGA). TGA
CC (when activated by calcium ions) is a gelling agent, useful in the
CC prodn. of foodstuffs, e.g. yoghurt, jelly, cheese and fishpaste.
SQ Sequence 3394 BP; 1024 A; 726 C; 778 G; 866 T;

Query Match 72.0%; Score 14.4; DB 1; Length 3394;
Best Local Similarity 93.8%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       2 CGCGTCGGTATTCCGA 17
           || ||||||||||||||
Db    1398 CGGGTCGGTATTCCGA 1383
```

Search completed: June 22, 2000, 15:12:40
Job time: 18086 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:57:56 ; Search time 8627.09 Seconds
(without alignments)
-2.255 Million cell updates/sec

Title: US-09-362-485-12
Perfect score: 20
Sequence: 1 GCGCGTCGGTATTCCGACCG 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20	100.0	20	5	A87615	A87615 Sequence 12
2	20	100.0	1194	2	MTU92472	U92472 Mycobacteri
3	20	100.0	1208	5	A89752	A89752 Sequence 9
4	20	100.0	1209	5	A87612	A87612 Sequence 9
5	20	100.0	1228	5	A87608	A87608 Sequence 5
6	20	100.0	1228	5	A89748	A89748 Sequence 5
7	20	100.0	1229	5	A87610	A87610 Sequence 7
8	20	100.0	1229	5	A89750	A89750 Sequence 7
9	20	100.0	1235	5	A87606	A87606 Sequence 3
10	20	100.0	1235	5	A87609	A87609 Sequence 6
11	20	100.0	1235	5	A87611	A87611 Sequence 8
12	20	100.0	1235	5	A89746	A89746 Sequence 3
13	20	100.0	1235	5	A89749	A89749 Sequence 6
14	20	100.0	1235	5	A89751	A89751 Sequence 8
15	20	100.0	1236	5	A87613	A87613 Sequence 10
16	20	100.0	1236	5	A89753	A89753 Sequence 10
17	20	100.0	1237	5	A87607	A87607 Sequence 4
18	20	100.0	1237	5	A89747	A89747 Sequence 4
19	20	100.0	1245	5	A87605	A87605 Sequence 2
20	20	100.0	1245	5	A89745	A89745 Sequence 2
21	20	100.0	1260	5	A87604	A87604 Sequence 1
22	20	100.0	1260	5	A89744	A89744 Sequence 1
23	20	100.0	2584	1	MTALADH	X63069 M.tuberculo
24	20	100.0	56414	1	MTV002	AL008967 Mycobacte
25	16.8	84.0	1363	35	MBPAFF1	AF006820 Pseudococ
c 26	16.4	82.0	38908	34	CEF37H8	Z81534 Caenorhabdi
27	16	80.0	18	5	A87614	A87614 Sequence 11
c 28	15.8	79.0	328	3	SSPMB35	X51562 S. scrofa A
29	15.8	79.0	1363	35	AF007263	AF007263 Pseudococ
30	15.8	79.0	1364	35	AF007265	AF007265 Pseudococ
31	15.8	79.0	1436	35	AF007264	AF007264 Pseudococ
32	15.8	79.0	36310	1	MLCB4	AL023514 Mycobacte
33	15.8	79.0	134226	16	IH1CG	M75136 Ictalurid h
34	15.2	76.0	534	8	AF042401	AF042401 Holterman
35	15.2	76.0	550	8	AF042451	AF042451 Tremella
36	15.2	76.0	567	8	OAU72601	U72601 Orbilia aln
37	15.2	76.0	567	8	OSU72597	U72597 Orbilia sp.
38	15.2	76.0	568	8	ODU72593	U72593 Orbilia del
39	15.2	76.0	594	8	ODU72595	U72595 Orbilia del
40	15.2	76.0	603	8	OAU72600	U72600 Orbilia aln
41	15.2	76.0	804	24	E09665	E09665 DNA encodin
42	15.2	76.0	804	24	E09666	E09666 DNA encodin
43	15.2	76.0	807	24	E09667	E09667 DNA encodin
44	15.2	76.0	1400	2	RGU67154	U67154 Rubrivivax
45	15.2	76.0	1400	2	RGU67155	U67155 Rubrivivax

ALIGNMENTS

```
RESULT   1
A87615
LOCUS       A87615         20 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 12 from Patent WO9836089.
ACCESSION   A87615
VERSION     A87615.1  GI:6736255
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      7 c      7 g      4 t
ORIGIN
```

Query Match 100.0%; Score 20; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db      1 GCGCGTCGGTATTCCGACCG 20
```

```
RESULT   2
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN
```

Query Match 100.0%; Score 20; DB 2; Length 1194;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82
```

```
RESULT   3
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN
```

Query Match 100.0%; Score 20; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82
```

```
RESULT   4
A87612
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 7.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT   5
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT   6
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT   7
A87610
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCG TATTCCGACCG 20
          |||||||| ||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT   8
A89750
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT   9
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
```

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  10
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  11
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  12
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  13
A89749
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  14
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82


RESULT  15
A87613
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGTCGGTATTCCGACCG 20
          ||||||||||||||||||||
Db     63 GCGCGTCGGTATTCCGACCG 82
```

Search completed: June 22, 2000, 14:57:58
Job time: 17893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:09 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-11
Perfect score: 18
Sequence: 1 ATGCGCGTCGGTATTCCG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	15.4	85.6	480	81	AJ282503	AJ282503 4A3A-AAY-
c	2	15	83.3	836	102	AQ163541	AQ163541 nbxb0007O
	3	14.8	82.2	158	37	AA716586	AA716586 zg66c12.s
c	4	14.8	82.2	638	64	AW065578	AW065578 614060D05
	5	14.4	80.0	166	91	AQ845925	AQ845925 LMAJFV1_1
c	6	14.4	80.0	424	46	AI399024	AI399024 NCW10B4T3
c	7	14.4	80.0	447	91	AQ882139	AQ882139 HS_5342_A
c	8	14.4	80.0	617	103	AQ194643	AQ194643 RPCI11-45
c	9	14.4	80.0	691	84	B20846	B20846 T29O3-T7 TA
	10	13.8	76.7	194	46	AI438741	AI438741 SWOv3MCAM
c	11	13.8	76.7	288	59	AV078417	AV078417 AV078417
c	12	13.8	76.7	300	62	AV183949	AV183949 AV183949
	13	13.8	76.7	319	79	AW257703	AW257703 SWOv3MCAM
	14	13.8	76.7	330	48	AI618873	AI618873 SWOv3MCAM
c	15	13.8	76.7	381	39	AA872699	AA872699 oi06d08.s
	16	13.8	76.7	387	79	AW288294	AW288294 SWOv3MCAM
c	17	13.8	76.7	406	38	AA788267	AA788267 r6c08a1.r
c	18	13.8	76.7	407	106	AQ394262	AQ394262 CITBI-E1-
c	19	13.8	76.7	412	109	AQ641999	AQ641999 RPCI93-Ec
c	20	13.8	76.7	452	31	AA291478	AA291478 zt40a08.s
c	21	13.8	76.7	528	45	AI387498	AI387498 GH18043.5
c	22	13.8	76.7	528	46	AI404854	AI404854 GH24695.5
c	23	13.8	76.7	532	45	AI387269	AI387269 GH17750.5
c	24	13.8	76.7	535	46	AI403382	AI403382 GH22808.5
	25	13.8	76.7	574	109	AQ641850	AQ641850 RPCI93-Dp
c	26	13.8	76.7	586	46	AI405354	AI405354 GH25328.5
c	27	13.8	76.7	587	82	FR0020269	AL013154 F.rubripe
c	28	13.8	76.7	602	105	AQ288318	AQ288318 nbxb0032L
	29	13.8	76.7	605	102	AQ160519	AQ160519 mgxb0006E
	30	13.8	76.7	622	36	AA618749	AA618749 SWOv3MCA1
	31	13.8	76.7	635	36	AA618752	AA618752 SWOv3MCA1
c	32	13.8	76.7	649	109	AQ660757	AQ660757 Sheared D
	33	13.8	76.7	652	109	AQ606097	AQ606097 hS_5392_A
c	34	13.8	76.7	672	84	B07838	B07838 8536RI01F72
c	35	13.8	76.7	673	109	AQ643946	AQ643946 RPCI93-Ec
	36	13.8	76.7	720	109	AQ654402	AQ654402 Sheared D
c	37	13.8	76.7	757	105	AQ291949	AQ291949 nbxb0040M
	38	13.8	76.7	801	91	AQ868420	AQ868420 nbeb0027B
c	39	13.8	76.7	816	91	AQ856530	AQ856530 nbeb0003J
	40	13.8	76.7	877	103	AQ255551	AQ255551 mgxb0014D
	41	13.4	74.4	121	44	AI324176	AI324176 mh80b05.x
	42	13.4	74.4	348	82	FR0033318	AL029685 Fugu rubr
c	43	13.4	74.4	358	37	T42645	T42645 5908 Lambda
	44	13.4	74.4	379	63	AI966107	AI966107 sc27e06.y
c	45	13.4	74.4	407	34	AA472790	AA472790 vd58a05.r

ALIGNMENTS

RESULT 1
AJ282503
LOCUS AJ282503 480 bp mRNA EST 07-FEB-2000
DEFINITION 4A3A-AAY-F-11-R Anopheles gambiae immune competent 4A3A Anopheles gambiae cDNA clone 4A3A-AAY-F-11, mRNA sequence.
ACCESSION AJ282503
VERSION AJ282503.1 GI:6930382
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anopheles.
REFERENCE 1 (bases 1 to 480)
AUTHORS Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.
TITLE Anopheles gambiae pilot gene discovery project: Identification of novel mosquito innate immunity genes from ESTs generated from immune competent cell lines
JOURNAL Unpublished (2000)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5409836.
Contact: Dimopoulos G
Fotis C. Kafatos laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES Location/Qualifiers
source 1. .480
/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-AAY-F-11"
/clone_lib="Anopheles gambiae immune competent 4A3A"
/cell_line="immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT 120 a 125 c 144 g 91 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 81; Length 480;
Best Local Similarity 94.1%; Pred. No. 74;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 TGCGCGTCGGTATTCCG 18
          ||||||||||||||| ||
Db    407 TGCGCGTCGGTATTACG 423
```

RESULT 2
AQ163541/c
LOCUS AQ163541 836 bp DNA GSS 12-SEP-1998
DEFINITION nbxb0007O19r CUGI Rice BAC Library Oryza sativa genomic clone nbxb0007O19r, genomic survey sequence.
ACCESSION AQ163541
VERSION AQ163541.1 GI:3553167
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 836)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 226.
FEATURES Location/Qualifiers
source 1. .836
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbxb0007O19r"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:

HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 174 a 186 c 199 g 275 t 2 others
ORIGIN

Query Match 83.3%; Score 15; DB 102; Length 836;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 CGCGTCGGTATTCCG 18
||||||||||||||||
Db 258 CGCGTCGGTATTCCG 244

RESULT 3
AA716586
LOCUS AA716586 158 bp mRNA EST 29-DEC-1997
DEFINITION zg66c12.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:398326 3' similar to gb:L04286 NK-TUMOR RECOGNITION PROTEIN (HUMAN);, mRNA sequence.
ACCESSION AA716586
VERSION AA716586.1 GI:2728860
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 158)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:811235.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .158
/organism="Homo sapiens"
/db_xref="GDB:1306149"
/db_xref="taxon:9606"
/clone="IMAGE:398326"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."
BASE COUNT 49 a 33 c 30 g 46 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 37; Length 158;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
|||||| ||||||||| |
Db 54 ATGCGCATCGGTATTCAG 71

RESULT 4
AW065578/c
LOCUS AW065578 638 bp mRNA EST 12-OCT-1999
DEFINITION 614060D05.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA, mRNA sequence.
ACCESSION AW065578
VERSION AW065578.1 GI:6020770
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE 1 (bases 1 to 638)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133420.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614060 row: D column: 05.
FEATURES Location/Qualifiers
source 1. .638
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="XLOLR"
/note="Organ: root; Vector: pBlueScriptII SK+; Site_1: EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot Lab (LM)"
BASE COUNT 194 a 118 c 157 g 168 t 1 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 64; Length 638;
Best Local Similarity 88.9%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18

```
          |||||| |||||||| ||
Db    320 ATGCGCTTCGGTATTGCG 303


RESULT   5
AQ845925
LOCUS       AQ845925      166 bp    DNA             GSS       15-OCT-1999
DEFINITION  LMAJFV1_lm11c05.y1 Leishmania major FV1 random genomic library
            Leishmania major genomic clone LMAJFV1_lm11c05 5' similar to
            contains element V1-ch1_type_II.1 V1-ch1_type_I leishmania
            repetitive element ;, genomic survey sequence.
ACCESSION   AQ845925
VERSION     AQ845925.1  GI:6050573
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 166)
  AUTHORS   Akopyants,N.S., Beverley,S.M., Clifton,S., Marra,M., Hillier,L.,
            Chinwalla,A., Pape,D., Martin,J., Wylie,T., Blistain,A.,
            Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko,I.,
            Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
            Harvey,N., McCann,R., Tsagareishvili,R., Williams,T., Jackson,Y.,
            Bowers,Y., Swaller,T., Waterston,R. and Wilson,R.
  TITLE     WashU Leishmania Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Akopyants, NS / Beverley, SM
            WashU Leishmania Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library construction: Natalia S. Akopyants, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             If using this information please cite:
             N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
            Friedlin strain V1 genome by shotgun sequencing' and the Washington
            University Genome Sequencing Center For information on obtaining
            clone material please contact: Natalia S. Akopyants Ph.D.
            (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
            (beverley@borcim.wustl.edu)
            Seq primer: -40RP from Gibco
            Class: shotgun
            High quality sequence stop: 146.
FEATURES              Location/Qualifiers
     source           1. .166
                      /organism="Leishmania major"
                      /strain="Friedlin strain V1"
                      /db_xref="taxon:5664"
                      /clone="LMAJFV1_lm11c05"
                      /clone_lib="Leishmania major FV1 random genomic library"
                      /lab_host="TOP10 (Invitrogen)"
                      /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
                      Genomic DNA was isolated from stationary phase cells. For
                      this library, DNA was sheared to give a tight size
                      distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
                      polymerase, dephosphorylated with Shrimp Alkaline
                      Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT       42 a     58 c     46 g     20 t
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 91;  Length 166;
  Best Local Similarity   93.8%;  Pred. No. 2.1e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TGCGCGTCGGTATTCC 17
           |||||||||||||  ||
Db     104 TGCGCGTCGGTATGCC 119


RESULT   6
AI399024/c
LOCUS       AI399024      424 bp    mRNA            EST       08-FEB-1999
DEFINITION  NCW10B4T3 Westergaards Neurospora crassa cDNA clone W10B4 5', mRNA
            sequence.
ACCESSION   AI399024
VERSION     AI399024.1  GI:4242109
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;
            Neurospora.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Erre· ,A., Fleharty,M., Gorman,M., Judson,K.,
            Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
            Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
            and Natvig,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949109.
            Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Neurospora crassa"
                      /strain="74-OR23-IV A (FGSC 2489)"
                      /db_xref="taxon:5141"
                      /clone="W10B4"
                      /clone_lib="Westergaards"
                      /sex="Mating type A"
                      /tissue_type="Unfertilized sexual tissue"
                      /dev_stage="Unfertilized sexual stage"
                      /lab_host="E. coli"
                      /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                      XhoI; Westergaard's medium (Nitrogen limiting). Floating
                      mycelial mats grown at 25C for 36 hours. cDNA
                      directionally cloned into pBluescript SK(-) using the
                      Uni-ZAP XR vector system (Stratagene, La Jolla, CA)."
BASE COUNT      106 a    114 c    101 g     96 t      7 others
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 46;  Length 424;
  Best Local Similarity   93.8%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GCGCGTCGGTATTCCG 18
           |||| ||||||||||||
Db     147 GCGCCTCGGTATTCCG 132


RESULT   7
AQ882139/c
LOCUS       AQ882139      447 bp    DNA             GSS       09-NOV-1999
DEFINITION  HS_5342_A2_F05_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9110 Col=10 Row=K, genomic survey sequence.
ACCESSION   AQ882139
VERSION     AQ882139.1  GI:6313606
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
```

```
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Feb 19, 1999 this sequence version replaced gi:4143280.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9110  row: K  column: 10
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=9110 Col=10 Row=K"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      151 a     89 c     72 g    133 t      2 others
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 91;  Length 447;
  Best Local Similarity   93.8%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTC 16
          ||||||| ||||||||
Db     56 ATGCGCGGCGGTATTC 41


RESULT   8
AQ194643/c
LOCUS       AQ194643      617 bp    DNA             GSS       20-APR-1999
DEFINITION  RPCI11-45F9.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-45F9,
            genomic survey sequence.
ACCESSION   AQ194643
VERSION     AQ194643.1  GI:3606255
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-45F9.TK
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Homo sapiens"
                     /db_xref="GDB:7517024"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-45F9"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      183 a    121 c    148 g    165 t
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 103;  Length 617;
  Best Local Similarity   93.8%;  Pred. No. 2.9e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TGCGCGTCGGTATTCC 17
          ||| ||||||||||||
Db    457 TGCACGTCGGTATTCC 442


RESULT   9
B20846/c
LOCUS       B20846        691 bp    DNA             GSS       16-SEP-1997
DEFINITION  T29O3-T7 TAMU Arabidopsis thaliana genomic clone T29O3, genomic
            survey sequence.
ACCESSION   B20846
VERSION     B20846.1  GI:2395900
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T29O3-Sp6
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 69
            High quality sequence stop: 151.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T29O3"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
BASE COUNT      83 a    311 c     58 g    236 t      3 others
ORIGIN
```

```
  Query Match             80.0%;  Score 14.4;  DB 84;  Length 691;
  Best Local Similarity   93.8%;  Pred. No. 3e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 GCGCGTCGGTATTCCG 18
          |||| ||||||||||||
Db    106 GCGCATCGGTATTCCG 91


RESULT  10
AI438741
LOCUS       AI438741      194 bp    mRNA            EST       09-MAR-1999
DEFINITION  SWOv3MCAM22C04SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWOv3MCAM22C04 5',
            mRNA sequence.
ACCESSION   AI438741
VERSION     AI438741.1  GI:4298294
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spiruida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 194)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188031.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .194
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWOv3MCAM22C04"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
                     /dev_stage="molting L3"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Filarial nematode parasite of humans. Third-stage
                     larvae, L3,  were isolated from infected black flies in
                     Cameroon (forest strain). The L3 were cultured in 20% FCS
                     in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
                     culture. L3 of O. volvulus molt to fourth-stage larvae by
                     day 5 in culture. mRNA was isolated from approximately
                     6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3
                     in culture, and converted to double-stranded cDNA using
                     reverse transcriptase and oligo(dT) followed by RNase H
                     and DNA pol I. The library was constructed in the lambda
                     Uni-Zap XR vector and has 1 x 10E6 independent
                     recombinants and the average insert size is ~1200 bp. The
                     library was constructed by Sara Lustigman and Michelle
                     Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
                     The library is available from Dr. Sara Lustigman (email:
                     slustigm@nybc.org)."
BASE COUNT       51 a     36 c     51 g     55 t      1 others
ORIGIN


  Query Match             76.7%;  Score 13.8;  DB 46;  Length 194;
  Best Local Similarity   88.2%;  Pred. No. 4.8e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 TGCGCGTCGGTATTCCG 18
          |||||||||||| || |
Db     21 TGCGCGTCGGTAATCGG 37


RESULT  11
AV078417/c
LOCUS       AV078417      288 bp    mRNA            EST       24-JUN-1999
DEFINITION  AV078417 Mus musculus stomach C57BL/6J adult Mus musculus cDNA
            clone 2210404B02, mRNA sequence.
ACCESSION   AV078417
VERSION     AV078417.1  GI:5198245
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187696.
            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1 1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2210404B02"
                     /clone_lib="Mus musculus stomach C57BL/6J adult"
                     /sex="male"
                     /tissue_type="stomach"
                     /dev_stage="adult"
BASE COUNT       77 a     55 c     49 g    107 t
ORIGIN


  Query Match             76.7%;  Score 13.8;  DB 59;  Length 288;
  Best Local Similarity   88.2%;  Pred. No. 5.3e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCC 17
          ||||| ||||| |||||
Db     71 ATGCGGGTCGGGATTCC 55


RESULT  12
AV183949/c
LOCUS       AV183949      300 bp    mRNA            EST       21-JUL-1999
DEFINITION  AV183949 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk658c10 3', mRNA
            sequence.
ACCESSION   AV183949
```

```
VERSION     AV183949.1  GI:5563850
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
COMMENT     On Feb 18, 1999 this sequence version replaced gi:4297658.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk658c10"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT      114 a     38 c     72 g     76 t
ORIGIN


  Query Match           76.7%;  Score 13.8;  DB 62;  Length 300;
  Best Local Similarity  88.2%;  Pred. No. 5.3e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ATGCGCGTCGGTATTCC 17
           |||||||||| |||||
Db     283 ATGCGCGTCGCAATTCC 267


RESULT  13
AW257703
LOCUS       AW257703      319 bp    mRNA            EST       20-DEC-1999
DEFINITION  SWOv3MCAM30H04SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWOv3MCAM30H04 5',
            mRNA sequence.
ACCESSION   AW257703
VERSION     AW257703.1  GI:6605960
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3136799.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWOv3MCAM30H04"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
                     /dev_stage="molting L3"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Filarial nematode parasite of humans. Third-stage
                     larvae, L3,  were isolated from infected black flies in
                     Cameroon (forest strain). The L3 were cultured in 20% FCS
                     in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
                     culture. L3 of O. volvulus molt to fourth-stage larvae by
                     day 5 in culture. mRNA was isolated from approximately
                     6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3
                     in culture, and converted to double-stranded cDNA using
                     reverse transcriptase and oligo(dT) followed by RNase H
                     and DNA pol I. The library was constructed in the lambda
                     Uni-Zap XR vector and has 1 x 10E6 independent
                     recombinants and the average insert size is ~1200 bp. The
                     library was constructed by Sara Lustigman and Michelle
                     Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
                     The library is available from Dr. Sara Lustigman (email:
                     slustigm@nybc.org)."
BASE COUNT      87 a     71 c     85 g     76 t
ORIGIN


  Query Match           76.7%;  Score 13.8;  DB 79;  Length 319;
  Best Local Similarity  88.2%;  Pred. No. 5.4e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TGCGCGTCGGTATTCCG 18
           |||||||||||| || |
Db      19 TGCGCGTCGGTAATCGG 35


RESULT  14
AI618873
LOCUS       AI618873      330 bp    mRNA            EST       21-APR-1999
DEFINITION  SWOv3MCAM27D11SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWOv3MCAM27D11 5',
            mRNA sequence.
ACCESSION   AI618873
VERSION     AI618873.1  GI:4627999
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948183.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWOv3MCAM27D11"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
```

/dev_stage="molting L3"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigm@nybc.org)."

BASE COUNT 93 a 70 c 88 g 79 t
ORIGIN

Query Match 76.7%; Score 13.8; DB 48; Length 330;
Best Local Similarity 88.2%; Pred. No. 5.5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTCCG 18
||||||||||||| || |
Db 34 TGCGCGTCGGTAATCGG 50

RESULT 15
AA872699/c
LOCUS AA872699 381 bp mRNA EST 14-APR-1998
DEFINITION oi06d08.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1475727 3', mRNA sequence.
ACCESSION AA872699
VERSION AA872699.1 GI:2968139
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 381)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405114.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium, LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1111 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 362.
FEATURES Location/Qualifiers
source 1. .381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1475727"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 103 a 100 c 85 g 93 t
ORIGIN

Query Match 76.7%; Score 13.8; DB 39; Length 381;
Best Local Similarity 88.2%; Pred. No. 5.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCC 17
|||||| |||| |||||
Db 331 ATGCGCATCGGCATTCC 315

Search completed: June 22, 2000, 12:08:13
Job time: 10407 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:26 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-11
Perfect score: 18
Sequence: 1 ATGCGCGTCGGTATTCCG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	13.8	76.7	15239	2	US-08-390-878-17	Sequence 17, Appl
c	2	13.4	74.4	2310	2	US-08-525-654A-2	Sequence 2, Appli
c	3	13.4	74.4	2313	2	US-08-525-654A-4	Sequence 4, Appli
c	4	13.4	74.4	3394	2	US-08-525-654A-136	Sequence 136, App
c	5	13.4	74.4	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	6	13.4	74.4	28958	1	US-08-456-837-6	Sequence 6, Appli
c	7	13.4	74.4	28958	1	US-08-457-342-6	Sequence 6, Appli
c	8	13.4	74.4	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	9	13.4	74.4	28958	2	US-08-458-076A-6	Sequence 6, Appli
c	10	13.4	74.4	28958	2	US-08-764-233A-4	Sequence 4, Appli
c	11	13.4	74.4	28958	2	US-08-457-335A-6	Sequence 6, Appli
c	12	13.4	74.4	28958	2	US-08-729-214-6	Sequence 6, Appli
c	13	13.4	74.4	49377	2	US-08-764-233A-1	Sequence 1, Appli
	14	13.2	73.3	1320	1	US-08-264-534-8	Sequence 8, Appli
	15	13.2	73.3	1320	2	US-08-083-590A-3	Sequence 3, Appli
	16	13.2	73.3	1320	2	US-08-465-500-8	Sequence 8, Appli
	17	13.2	73.3	1320	3	US-08-346-126-8	Sequence 8, Appli
	18	13.2	73.3	1320	3	US-08-346-128-8	Sequence 8, Appli
	19	13.2	73.3	5561	3	US-08-400-159-1	Sequence 1, Appli
	20	13.2	73.3	34303	4	US-08-735-609-4	Sequence 4, Appli
	21	13.2	73.3	34382	3	US-08-374-483-6	Sequence 6, Appli
	22	13.2	73.3	35935	4	US-08-735-609-1	Sequence 1, Appli
	23	13.2	73.3	35935	5	US-08-379-452-43	Sequence 43, Appl
c	24	12.8	71.1	660	3	US-08-401-068-5	Sequence 5, Appli
c	25	12.8	71.1	660	3	US-08-846-338-5	Sequence 5, Appli
	26	12.8	71.1	2452	3	US-08-524-051-1	Sequence 1, Appli
c	27	12.8	71.1	6895	2	US-08-570-311-9	Sequence 9, Appli
c	28	12.8	71.1	6895	3	US-08-353-485-9	Sequence 9, Appli
	29	12.4	68.9	795	5	US-09-053-197A-6	Sequence 6, Appli
	30	12.4	68.9	314'	4	US-08-658-665-66	Sequence 66, Appl
	31	12.4	68.9	3531	2	US-08-629-600-1	Sequence 1, Appli
c	32	12.4	68.9	3567	4	US-08-658-665-69	Sequence 69, Appl
c	33	12.4	68.9	4075	4	US-08-658-665-67	Sequence 67, Appl
c	34	12.4	68.9	4909	4	US-08-658-665-68	Sequence 68, Appl
c	35	12.4	68.9	6749	4	US-08-658-665-71	Sequence 71, Appl
	36	12.4	68.9	6749	4	US-08-658-665-177	Sequence 177, App
	37	12.4	68.9	11601	3	US-08-222-617A-3	Sequence 3, Appli
	38	12.4	68.9	11601	3	US-08-222-617A-24	Sequence 24, Appl
	39	12.4	68.9	22108	5	US-09-053-197A-3	Sequence 3, Appli
c	40	12.2	67.8	1092	1	US-08-400-422-1	Sequence 1, Appli
c	41	12.2	67.8	1301	4	US-08-641-314C-1	Sequence 1, Appli
c	42	12.2	67.8	1350	3	US-08-686-350-1	Sequence 1, Appli
c	43	12.2	67.8	1504	3	US-08-878-989-10	Sequence 10, Appl
c	44	12.2	67.8	1576	4	US-09-197-378-1	Sequence 1, Appli
c	45	12.2	67.8	16	1	US-07-959-941-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-390-878-17/c
; Sequence 17, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-17

Query Match 76.7%; Score 13.8; DB 2; Length 15239;
Best Local Similarity 88.2%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 TGCGCGTCGGTATTCCG 18
           |||||| |||||| |||
Db    1923 TGCGCGCCGGTATACCG 1907
```

```
RESULT   2
US-08-525-654A-2/c
; Sequence 2, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
;    TITLE OF INVENTION:  TRANSGLUTAMINASE ORIGINATED FROM
;    TITLE OF INVENTION:  CRASSOSTREA GIGAS
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,654A
;      FILING DATE:  28-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6/8283
;      FILING DATE:  28-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7/3876
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-760-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2310 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Crassostrea gigas
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2310
US-08-525-654A-2


  Query Match             74.4%;  Score 13.4;  DB 2;  Length 2310;
  Best Local Similarity   93.3%;  Pred. No. 50;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 CGCGTCGGTATTCCG 18
           || ||||||||||||
Db    1091 CGGGTCGGTATTCCG 1077


RESULT   3
US-08-525-654A-4/c
; Sequence 4, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
;    TITLE OF INVENTION:  TRANSGLUTAMINASE ORIGINATED FROM
;    TITLE OF INVENTION:  CRASSOSTREA GIGAS
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,654A
;      FILING DATE:  28-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6/8283
;      FILING DATE:  28-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7/3876
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-760-0 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2313 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Crassostrea gigas
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2054
US-08-525-654A-4


  Query Match             74.4%;  Score 13.4;  DB 2;  Length 2313;
  Best Local Similarity   93.3%;  Pred. No. 50;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 CGCGTCGGTATTCCG 18
           || ||||||||||||
Db    1094 CGGGTCGGTATTCCG 1080


RESULT   4
US-08-525-654A-136/c
; Sequence 136, Application US/08525654A
; Patent No. 5736356
;  GENERAL INFORMATION:
;    APPLICANT:  SANO, KOHICHIRO
;    APPLICANT:  KUMAZAWA, YOSHIYUKI
;    APPLICANT:  YASEUDA, HISASHI
;    APPLICANT:  SEGURO, KATSUYA
;    APPLICANT:  MOTOKI, MASAO
```

; TITLE OF INVENTION: TRANSGLUTAMINASE ORIGINATED FROM
; TITLE OF INVENTION: CRASSOSTREA GIGAS
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,654A
; FILING DATE: 28-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6/8283
; FILING DATE: 28-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7/3876
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-760-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Crassostrea gigas
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 305..2617
US-08-525-654A-136

Query Match 74.4%; Score 13.4; DB 2; Length 3394;
Best Local Similarity 93.3%; Pred. No. 50;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGCGTCGGTATTCCG 18
|| ||||||||||||
Db 1398 CGGGTCGGTATTCCG 1384

RESULT 5
US-08-258-261B-6/c
; Sequence 6, Application US/08258261B
; Patent No. 5637949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the syntnesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match 74.4%; Score 13.4; DB 1; Length 28958;
Best Local Similarity 93.3%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCC 17
|||||||| ||||||
Db 21961 GCGCGTCGATATTCC 21947

RESULT 6
US-08-456-837-6/c
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 28958;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GCGCGTCGGTATTCC 17
           ||||||||  ||||||
Db   21961 GCGCGTCGATATTCC 21947


RESULT   7
US-08-457-342-6/c
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 28958;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GCGCGTCGGTATTCC 17
           ||||||||  ||||||
Db   21961 GCGCGTCGATATTCC 21947


RESULT   8
US-08-457-646A-6/c
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 28958;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GCGCGTCGGTATTCC 17
           ||||||||| ||||||
Db   21961 GCGCGTCGATATTCC 21947


RESULT   9
US-08-458-076A-6/c
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6


  Query Match             74.4%;  Score 13.4;  DB 2;  Length 28958;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 GCGCGTCGGTATTCC 17
           ||||||||| ||||||
Db   21961 GCGCGTCGATATTCC 21947


RESULT  10
US-08-764-233A-4/c
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  1059
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1
```

US-08-764-233A-4

Query Match 74.4%; Score 13.4; DB 2; Length 28958;
Best Local Similarity 93.3%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCC 17
||||||||| ||||||
Db 21961 GCGCGTCGATATTCC 21947

RESULT 11
US-08-457-335A-6/c
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8[illegible]89
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-335A-6

Query Match 74.4%; Score 13.4; DB 2; Length 28958;
Best Local Similarity 93.3%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCC 17
||||||||| ||||||
Db 21961 GCGCGTCGATATTCC 21947

RESULT 12
US-08-729-214-6/c
; Sequence 6, Application US/08729214
; Patent No. 5817502
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,214
; FILING DATE: TBA
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8[illegible]9
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-729-214-6

Query Match 74.4%; Score 13.4; DB 2; Length 28958;
Best Local Similarity 93.3%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCC 17
||||||||| ||||||
Db 21961 GCGCGTCGATATTCC 21947

RESULT 13
US-08-764-233A-1/c
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezanna

; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "SorR"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous
; OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "SorA"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs that
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product= "SorB"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS ge
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product= "Module 1 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product= "Module 2 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30881..35446
; OTHER INFORMATION: /product= "Module 3 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product= "Module 4 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40190..46318
; OTHER INFORMATION: /product= "Module 5 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "SorM"
; OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rappamicin."
US-08-764-233A-1

Query Match 74.4%; Score 13.4; DB 2; Length 49377;
Best Local Similarity 93.3%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCC 17
||||||||| ||||||
Db 38264 GCGCGTCGATATTCC 38250

RESULT 14
US-08-264-534-8
; Sequence 8, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,534
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,189
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-004

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION   442..1320
US-08-264-534-8


  Query Match             73.3%;  Score 13.2;  DB 1;  Length 1320;
  Best Local Similarity   83.3%;  Pred. No. 63;
  Matches   15;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGCGCGTCGGTATTCCG 18
           |||| || ||| ||||||
Db     798 ATGCACGACGGCATTCCG 815


RESULT  15
US-08-083-590A-3
; Sequence 3, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  442..1320
US-08-083-590A-3
```

```
  Query Match             73.3%;  Score 13.2;  DB 2;  Length 1320;
  Best Local Similarity   83.3%;  Pred. No. 63;
  Matches   15;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGCGCGTCGGTATTCCG 18
           |||| || ||| ||||||
Db     798 ATGCACGACGGCATTCCG 815


Search completed: June 22, 2000, 15:21:32
Job time: 18770 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:35 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-11
Perfect score: 18
Sequence: 1 ATGCGCGTCGGTATTCCG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49512	Mycobacterium sp.
2		18	100.0	18	1	V49610	AlaDH derived olig
3		18	100.0	1245	1	V49626	Mycobacterium tube
4		18	100.0	1260	1	V49510	Mycobacterium sp.
5		18	100.0	1260	1	V49625	Mycobacterium tube
6		16	88.9	20	1	V49513	Mycobacterium sp.
7		16	88.9	20	1	V49611	AlaDH derived olig
8		13.8	76.7	64	1	T92901	RWZ13 secondary st
c 9		13.8	76.7	1038	1	V84117	Pseudomonas aerugi
10		13.8	76.7	1323	1	X20617	Polynucleotide seq
11		13.8	76.7	1935	1	X13361	Enterococcus faeca
12		13.8	76.7	6426	1	X20538	Polynucleotide seq
c 13		13.8	76.7	15239	1	T33536	BCG deletion regio
14		13.4	74.4	898	1	V63918	Mycobacterium tube
15		13.4	74.4	1802	1	N60715	Sequence encoding
16		13.4	74.4	2360	1	Q65321	Vaccinia virus 7.5
17		13.4	74.4	2360	1	Q86798	Vaccinia virus vec
c 18		13.4	74.4	3394	1	Q99223	Japanese oyster tr
c 19		13.4	74.4	3811	1	X38291	E. coli secA DNA.
20		13.4	74.4	4987	1	Q65322	Vaccinia virus pro
21		13.4	74.4	4987	1	Q86799	Vaccinia virus vec
c 22		13.4	74.4	28598	1	T06769	Sorangium cellulos
c 23		13.4	74.4	28958	1	T89956	Sorangium cellulos
c 24		13.4	74.4	49377	1	V05287	The soraphen biosy
25		13.2	73.3	1320	1	Q30998	Serrate gene. Huma
26		13.2	73.3	1713	1	Q03827	Pyruvate oxidase (
27		13.2	73.3	1713	1	Q08597	Pyruvate oxidase (
28		13.2	73.3	4260	1	N81768	Sequence encoding
29		13.2	73.3	4380	1	N80222	Sequence of Mycoba
30		13.2	73.3	4380	1	V05708	Mycobacterium tube
c 31		13.2	73.3	4800	1	X05350	Phenol hydroxylase
c 32		13.2	73.3	4800	1	X01590	Pseudomonas putida
c 33		13.2	73.3	5420	1	X20601	Polynucleotide seq
34		13.2	73.3	5561	1	Q43910	Sequence encoding
35		13.2	73.3	12145	1	V35743	Klebsiella pneumon
36		13.2	73.3	32026	1	T60559	Recombinant adenov
37		13.2	73.3	34303	1	V07261	Adenoviral vector
38		13.2	73.3	34382	1	X15627	Recombinant adenov
39		13.2	73.3	34427	1	X07371	Adenovirus vector
40		13.2	73.3	35000	1	T60557	Recombinant cis-ac
41		13.2	73.3	35408	1	T59272	Recombinant adenov
42		13.2	73.3	35935	1	V07258	Adenovirus 5 genom
43		13.2	73.3	36335	1	Q68003	Ad2/-ORF6/PGK-CFTR
c 44		13.2	73.3	110000	1	V30458_1	Continuation (2 of
c 45		13.2	73.3	110000	1	V30459_1	Continuation (2 of

ALIGNMENTS

RESULT 1
V49512
ID V49512 standard; DNA; 18 BP.
AC V49512;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F1.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 2 A; 5 C; 6 G; 5 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
||||||||||||||||||
Db 1 ATGCGCGTCGGTATTCCG 18

RESULT 2
V49610
ID V49610 standard; DNA; 18 BP.
AC V49610;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F1.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 2 A; 5 C; 6 G; 5 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
||||||||||||||||||
Db 1 ATGCGCGTCGGTATTCCG 18

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
||||||||||||||||||
Db 61 ATGCGCGTCGGTATTCCG 78

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
||||||||||||||||||
Db 76 ATGCGCGTCGGTATTCCG 93

RESULT 5
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
||||||||||||||||||
Db 76 ATGCGCGTCGGTATTCCG 93

RESULT 6
V49513
ID V49513 standard; DNA; 20 BP.
AC V49513;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F1+.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 20 BP; 2 A; 7 C; 7 G; 4 T;

Query Match 88.9%; Score 16; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCCG 18
||||||||||||||||
Db 1 GCGCGTCGGTATTCCG 16

RESULT 7
V49611
ID V49611 standard; DNA; 20 BP.
AC V49611;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F1+.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 20 BP; 2 A; 7 C; 7 G; 4 T;

Query Match 88.9%; Score 16; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCGCGTCGGTATTCCG 18
||||||||||||||||
Db 1 GCGCGTCGGTATTCCG 16

RESULT 8
T92901
ID T92901 standard; RNA; 64 BP.
AC T92901;
DT 24-APR-1998 (first entry)
DE RWZ13 secondary structure substrate RNA.
KW RWZ2; linked binding site; HIV Rev protein; inhibitor; replication;
KW human immunodeficiency virus; Rev response element; RRE; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_binding 1. .16
FT /*tag= a
FT /note= "binds to positions 49 to 64"
FT misc_structure 16. .17
FT /*tag= b
FT /note= "forms a loop with positions 46 to 49"
FT misc_structure 18. .22
FT /*tag= c
FT /note= "forms a loop with positions 42 to 45"
FT stem_loop 22. .42
FT /*tag= d
FT misc_structure 42. .45
FT /*tag= e
FT /note= "forms a loop with positions 22 to 18"
FT misc_structure 46. .49
FT /*tag= f
FT /note= "forms a loop with positions 17 to 16"
FT misc_binding 49. .64
FT /*tag= g
FT /note= "binds to positions 16 to 1"
PN WO9739128-A2.
PD 23-OCT-1997.
PF 15-APR-1997; G01054.
PR 13-MAY-1996; US-017268.
PR 15-APR-1996; GB-007819.
PA (MEDI-) MEDICAL RES COUNCIL.
PA (THER-) THEREXSYS LTD.

PI Butler PJG, Craig RK, Irvine AS, Karn J, Zemmel RW;
DR WPI; 97-526464/48.
PT Nucleic acid with two linked binding sites for HIV Rev protein -
PT used to treat or prevent HIV infection and to screen for Rev binding
PT inhibitors
PS Disclosure; Fig 8; 107pp; English.
CC A novel isolated nucleic acid has been developed which has two
CC operatively linked binding sites for human immunodeficiency virus (HIV)
CC Rev protein, each site comprising a nucleation motif (NM) and an
CC oligomerisation motif (OM). The novel isolated nucleic acid binds Rev
CC protein monomers with a greater degree of co-operativity than wild-type
CC Rev response element (RRE). The present sequence represents an RNA
CC substrate molecule mentioned in the present invention. The novel
CC isolated nucleic acid is used to treat, or protect against, HIV
CC infection (it acts as a decoy to bind Rev, thus inhibiting HIV
CC replication). The novel isolated nucleic acid is also used to screen
CC for compounds (A) that inhibit binding of Rev to a nucleic acid
CC substrate, the identified compounds are potential anti-HIV agents. The
CC vectors may be used to deliver the novel isolated nucleic acid in vivo
CC or for production of it for subsequent recovery and administration.
CC The novel nucleic acid binds (i.e. sequester) Rev more efficiently than
CC RRE. The initial binding of Rev to the high affinity NM is not in itself
CC sufficient to initiate Rev oligomerisation, for which the bordering OM
CC are required.
SQ Sequence 64 BP; 11 A; 18 C; 21 G; 14 U;

Query Match 76.7%; Score 13.8; DB 1; Length 64;
Best Local Similarity 64.7%; Pred. No. 56;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTCCG 18
:| |||:||| |::|||
Db 41 UGGGCGUCGGAAUUCCG 57

RESULT 9
V84117/c
ID V84117 standard; DNA; 1038 BP.
AC V84117;
DT 15-MAR-1999 (first entry)
DE Pseudomonas aeruginosa heptosyl transferase II waaF gene.
KW WaaF gene; rfaF gene; lipopolysaccharide; infection; therapy;
KW diagnosis; vaccine; heptosyl transferase II; ss.
OS Pseudomonas aeruginosa strain PAO1.
PN WO9850557-A1.
PD 12-NOV-1998.
PF 01-MAY-1998; CA0395.
PR 09-MAY-1997; US-046149.
PR 02-MAY-1997; US-045418.
PA (UYGU-) UNIV GUELPH.
PI Burrows LL, De Kievit TR, Lam JS, Matewish M, Walsh A;
DR WPI; 99-034725/03.
DR P-PSDB; W88211.
PT Isolated P. aeruginosa waaC, waaG, waaF and waaP gene cluster -
PT useful in the diagnosis or treatment of P. aeruginosa infections
PS Claim 2; Fig 6; 61pp; English.
CC This is the nucleotide sequence of the waaF gene (rfaF gene) of
CC the waa gene cluster of Pseudomonas aeruginosa PAO1. It codes for
CC WaaF (see W82211), a heptosyl transferase II that adds the second
CC heptose residue onto the core oligosaccharide in the biosynthesis
CC of the the lipopolysaccharide inner core. The four waa genes
CC of P. aeruginosa (see V84116-19) are arranged contiguously in an
CC operon with the gene order waaF, waaC, waaG and waaP. The
CC functions of the encoded proteins (see W82210-13) were tested by
CC complementation of specific Salmonella typhimurium mutants, and
CC knockout mutations of the genes in P. aeruginosa. The waa nucleic
CC acids or proteins can be used to diagnose a bacterial, especially a
CC P. aeruginosa, infection in an animal. They can further be used to
CC screen for compounds that affect core lipopolysaccharide biosynthesis
CC or assembly. A claimed method of treating or preventing a bacterial
CC infection comprises administering an agent that inhibits a waa gene
CC to an animal. A claimed vaccine for treating a bacterial infection
CC includes one of WaaP, WaaC, WaaF and WaaG.
SQ Sequence 1038 BP; 148 A; 348 C; 364 G; 178 T;

Query Match 76.7%; Score 13.8; DB 1; Length 1038;
Best Local Similarity 88.2%; Pred. No. 57;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTCCG 18
|| ||| ||||||||||||
Db 313 TACGCTTCGGTATTCCG 297

RESULT 10
X20617
ID X20617 standard; DNA; 1323 BP.
AC X20617;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 732-733; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 1323 BP; 3[illegible] A; 308 C; 320 G; 336 T;

Query Match 76.7%; Score 13.8; DB 1; Length 1323;
Best Local Similarity 88.2%; Pred. No. 58;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCC 17
| ||||||| |||||||
Db 1282 AAGCGCGTCCGTATTCC 1298

RESULT 11
X13361
ID X13361 standard; DNA; 1935 BP.
AC X13361;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:424.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

PS Claim 1; Page 1646-1647; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 1935 BP; 699 A; 259 C; 334 G; 637 T;

Query Match 76.7%; Score 13.8; DB 1; Length 1935;
Best Local Similarity 88.2%; Pred. No. 58;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTCCG 18
||| ||| |||||||||
Db 228 TGCTCGTTGGTATTCCG 244

RESULT 12
X20538
ID X20538 standard; DNA; 6426 BP.
AC X20538;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 410-414; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 6426 BP; 1587 A; 2076 C; 1514 G; 1243 T;

Query Match 76.7%; Score 13.8; DB 1; Length 6426;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGCGCGTCGGTATTCCG 18
| ||||| |||| |||||
Db 6404 ANGCGCGCCGGTTTTCCG 6421

RESULT 13
T33536/c
ID T33536 standard; DNA; 15239 BP.
AC T33536;
DT 15-FEB-1998 (first entry)
DE BCG deletion region 2 and flanking sequences.
KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
OS Mycobacterium bovis strain BCG.
FH Key Location/Qualifiers
FT misc_feature 3382. .14071
FT /*tag= a
FT /note= "BCG delta 1 deletion region"
PN WO9625519-A1.
PD 22-AUG-1996.
PF 15-FEB-1996; U01938.
PR 17-FEB-1995; US-390878.
PA (PATH-) PATHOGENESIS CORP.
PI Mahairas GG, Stover CK;
DR WPI; 96-393419/39.
PT Detecting markers for avirulence in Mycobacterium - used in
PT production of vaccines against bacterial infection, and to detect
PT bacterial infection
PS Example 1; Fig 2; 66pp; English.
CC This DNA sequence comprises Mycobacterium bovis BCG deletion
CC sequence BCGdelta2. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 other
CC deletion regions (see T33535 and T33537) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absense of deletion junctions (see T33538-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T;

Query Match 76.7%; Score 13.8; DB 1; Length 15239;
Best Local Similarity 88.2%; Pred. No. 59;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTCCG 18
|||||| |||||| |||
Db 1923 TGCGCGCCGGTATACCG 1907

RESULT 14
V63918
ID V63918 standard; DNA; 898 BP.
AC V63918;
DT 21-JAN-1999 (first entry)
DE Mycobacterium tuberculosis antigen CFP20 encoding DNA.
KW Mycobacterium tuberculosis; antigen; vaccine; immunological;
KW immunogen; infection; ds.
OS Mycobacterium tuberculosis.
FH Key Location/Qualifiers
FT CDS 201. .698
FT /*tag= a
PN WO9844119-A1.
PD 08-OCT-1998.
PF 01-APR-1998; DK0132.
PR 05-JAN-1998; US-070488.
PR 02-APR-1997; DK-000376.
PR 18-APR-1997; US-044624.
PR 10-NOV-1997; DK-001277.
PA (STAT-) STATENS SERUM INST.
PI Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB,
PI Rosenkrands I, Weldingh K;
DR WPI; 98-542705/46.
DR P-PSDB; W72887.
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
PS Claim 23; Page 128-129; 163pp; English.
CC The present sequence encodes a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments

CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
SQ Sequence 898 BP; 160 A; 294 C; 299 G; 145 T;

Query Match 74.4%; Score 13.4; DB 1; Length 898;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CGCGTCGGTATTCCG 18
||||||||| ||||||
Db 509 CGCGTCGGCATTCCG 523

RESULT 15
N60715
ID N60715 standard; DNA; 1802 BP.
AC N60715;
DT 17-OCT-1991 (first entry)
DE Sequence encoding vaccina virus haemagglutinin gene.
KW HA; living vaccine; ss.
OS Vaccina virus.
FH Key Location/Qualifiers
FT cds 444. .1391
FT /*tag= a
PN J61146187-A.
PD 03-JUL-1986.
PF 21-DEC-1984; 268520.
PR 21-DEC-1984; JP-268520.
PA (KYOU) KYOTO DAIGAKU.
DR WPI; 86-215125/33.
DR P-PSDB; P61726.
PT Plasmid contg. vaccinia virus haemagglutinin (HA)-gene - used for
PT mfg. transformant virus for prodn. of living vaccines.
PS Disclosure; Page 474; 9pp; Japanese.
CC HA gene may be used for the production of transformant viruses for
CC living vaccines.
SQ Sequence 1802 BP; 643 A; 313 C; 313 G; 533 T;

Query Match 74.4%; Score 13.4; DB 1; Length 1802;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TGCGCGTCGGTATTC 16
||| ||||||||||
Db 927 TGCTCGTCGGTATTC 941

Search completed: June 22, 2000, 15:12:38
Job time: 18084 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:57:54 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-11
Perfect score: 18
Sequence: 1 ATGCGCGTCGGTATTCCG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87614	A87614 Sequence 11
2		18	100.0	1194	2	MTU92472	U92472 Mycobacteri
3		18	100.0	1208	5	A89752	A89752 Sequence 9
4		18	100.0	1209	5	A87612	A87612 Sequence 9
5		18	100.0	1228	5	A87608	A87608 Sequence 5
6		18	100.0	1228	5	A89748	A89748 Sequence 5
7		18	100.0	1229	5	A87610	A87610 Sequence 7
8		18	100.0	1229	5	A89750	A89750 Sequence 7
9		18	100.0	1235	5	A87606	A87606 Sequence 3
10		18	100.0	1235	5	A87609	A87609 Sequence 6
11		18	100.0	1235	5	A87611	A87611 Sequence 8
12		18	100.0	1235	5	A89746	A89746 Sequence 3
13		18	100.0	1235	5	A89749	A89749 Sequence 6
14		18	100.0	1235	5	A89751	A89751 Sequence 8
15		18	100.0	1236	5	A87613	A87613 Sequence 10
16		18	100.0	1236	5	A89753	A89753 Sequence 10
17		18	100.0	1237	5	A87607	A87607 Sequence 4
18		18	100.0	1237	5	A89747	A89747 Sequence 4
19		18	100.0	1245	5	A87605	A87605 Sequence 2
20		18	100.0	1245	5	A89745	A89745 Sequence 2
21		18	100.0	1260	5	A87604	A87604 Sequence 1
22		18	100.0	1260	5	A89744	A89744 Sequence 1
23		18	100.0	2584	1	MTALADH	X63069 M.tuberculo
24		18	100.0	56414	1	MTV002	AL008967 Mycobacte
25		16	88.9	20	5	A87615	A87615 Sequence 12
26	c	15.4	85.6	7991	2	PAU63816	U63816 Pseudomonas
27		15.4	85.6	184270	43	AC008782	AC008782 Homo sapi
28	c	14.8	82.2	11283	2	AF144422	AF144422 Salmonell
29		14.8	82.2	39800	1	MTCY10D7	Z79700 Mycobacteri
30	c	14.8	82.2	125727	55	AC007928	AC007928 Drosophil
31		14.8	82.2	155065	11	HSJ659I19	AL122003 Human DNA
32	c	14.8	82.2	172815	55	AC010920	AC010920 Drosophil
33		14.8	82.2	190289	16	MCU60315	U60315 Molluscum c
34	c	14.8	82.2	237119	43	AC017740	AC017740 Drosophil
35		14.4	80.0	326	8	PFITS2	U90809 Potentilla
36		14.4	80.0	768	4	PTSATF	X14379 Phylloscopu
37		14.4	80.0	1363	35	MBPAFF1	AF006820 Pseudococ
38		14.4	80.0	1927	1	PSENAPDOXB	M83950 P.putida na
39	c	14.4	80.0	1965	1	PSU18135	U18135 Pseudomonas
40		14.4	80.0	2059	7	CCATCOLE1	Y10421 C.curvatus
41	c	14.4	80.0	2829	2	AF218420	AF218420 Brucella
42		14.4	80.0	4355	1	PSENAPDOXA	M83949 P.putida na
43		14.4	80.0	9706	2	AF010471	AF010471 Pseudomon
44		14.4	80.0	14462	1	AB004059	AB004059 Pseudomon
45	c	14.4	80.0	14520	16	VACLEND	M22812 Vaccinia vi

ALIGNMENTS

```
RESULT   1
A87614
LOCUS       A87614         18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 11 from Patent WO9836089.
ACCESSION   A87614
VERSION     A87614.1  GI:6736254
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      5 c      6 g      5 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db      1 ATGCGCGTCGGTATTCCG 18


RESULT   2
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   3
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   4
A87612
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 1209;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   5
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   6
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   7
A87610
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   8
A89750
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT   9
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
```

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  10
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  11
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  12
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  13
A89749
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  14
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERI M MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78


RESULT  15
A87613
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1236;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGCGCGTCGGTATTCCG 18
          ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCG 78
```

```
Search completed: June 22, 2000, 14:57:56
Job time: 17891 sec
```

```
RESULT   1
A87620
LOCUS       A87620          21 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 17 from Patent WO9836089.
ACCESSION   A87620
VERSION     A87620.1  GI:6736260
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      8 c      6 g      3 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db      1 GACGTCGACCTACGCGCTGAC 21


RESULT   2
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT   3
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (D  ; SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT   4
A87612
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT       228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT   5
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT   6
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT   7
A87610
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT   8
A89750
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT   9
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
```

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5   Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT  10
A87609
LOCUS       A87609       1235 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT  11
A87611
LOCUS       A87611       1235 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT  12
A89746
LOCUS       A89746       1235 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN

  Query Match             100.0%;  Score 21;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001


RESULT  13
A89749
LOCUS       A89749       1235 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match             100.0%;  Score 21;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT  14
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    980 GACGTCGACCTACGCGCTGAC 1000


RESULT  15
A87613
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db    981 GACGTCGACCTACGCGCTGAC 1001
```

Search completed: June 22, 2000, 14:58:36
Job time: 17931 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:51 ; Search time 458.59 Seconds
(without alignments)
11.457 Million cell updates/sec

Title: US-09-362-485-17
Perfect score: 21
Sequence: 1 GACGTCGACCTACGCGCTGAC 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	21	1	V49518	Mycobacterium sp.
2		21	100.0	21	1	V49616	AlaDH derived olig
3		21	100.0	1245	1	V49626	Mycobacterium tube
4		21	100.0	1260	1	V49510	Mycobacterium sp.
5		21	100.0	1260	1	V49625	Mycobacterium tube
c	6	16.2	77.1	2538	1	Q92232	Brevibacterium fla
c	7	16.2	77.1	23666	1	Q10190	Cephalosporin anti
8		16.2	77.1	24379	1	T93095	Streptomyces freno
9		16.2	77.1	24379	1	V25925	Streptomyces roseo
10		15.2	72.4	358	1	Q37959	Streptomyces ghana
c	11	15.2	72.4	398	1	V20280	Probe (181) for mi
12		15.2	72.4	1023	1	V73507	E. chrysanthemi Hr
13		15.2	72.4	1125	1	N91423	Sequence of heat-r
14		15.2	72.4	1746	1	T92239	Mercuric ion reduc
15		15.2	72.4	1752	1	T92237	Mercuric ion reduc
16		15.2	72.4	1752	1	T92238	Mercuric ion reduc
17		15.2	72.4	1752	1	T92240	Mercuric ion reduc
18		15.2	72.4	1752	1	T92241	Mercuric ion reduc
19		15.2	72.4	2107	1	T11024	Urochloa panicoide
20		15.2	72.4	2141	1	T49313	Hypersensitive res
21		15.2	72.4	2141	1	V36427	Hypersensitive res
22		15.2	72.4	2141	1	V39972	Erwinia chrysanthe
23		15.2	72.4	2141	1	V54606	Erwinia chrysanthe
24		15.2	72.4	2141	1	V73494	E. chrysanthemi Hr
25		15.2	72.4	2141	1	V83988	DNA encoding a hyp
26		15.2	72.4	2220	1	T11022	Urochloa panicoide
27		15.2	72.4	2245	1	T11023	Urochloa panicoide
c	28	15.2	72.4	3606	1	Q13370	Fusaric acid resis
c	29	15.2	72.4	7076	1	Q20186	HlyA gene with ctx
c	30	15.2	72.4	7076	1	Q86906	Vibrio cholerae di
c	31	14.8	70.5	2294	1	T63246	Genomic sequence e
c	32	14.8	70.5	24379	1	T93095	Streptomyces freno
c	33	14.8	70.5	24379	1	V25925	Streptomyces roseo
34		14.6	69.5	866	1	X25148	Wheat Type I gluta
c	35	14.6	69.5	1254	1	Q73586	Fragment of the H.
36		14.6	69.5	1268	1	Q64202	snaA gene encoding
c	37	14.6	69.5	1632	1	Q96251	Fructosyltransfera
c	38	14.6	69.5	2177	1	T14174	A. niger strain M-
39		14.6	69.5	5392	1	Q64201	Sequence comprisin
40		14.6	69.5	15872	1	T68715	Streptomyces venez
c	41	14.6	69.5	16836	1	V52831	Acetobacter xylinu
c	42	14.4	68.6	1121	1	V64418	Mouse developing l
43		14.4	68.6	2802	1	T33521	Phosphoenolpyruvic
44		14.4	68.6	2802	1	T32996	Acetobacter hansen
45		14.2	67.6	267	1	T29632	Wild-type viral US

ALIGNMENTS

RESULT 1
V49518
ID V49518 standard; DNA; 21 BP.
AC V49518;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F7.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 21 BP; 4 A; 8 C; 6 G; 3 T;

Query Match 100.0%; Score 21; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACGTCGACCTACGCGCTGAC 21
||||||||||||||||||||||
Db 1 GACGTCGACCTACGCGCTGAC 21

RESULT 2
V49616
ID V49616 standard; DNA; 21 BP.
AC V49616;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F7.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 21 BP; 4 A; 8 C; 6 G; 3 T;

Query Match 100.0%; Score 21; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACGTCGACCTACGCGCTGAC 21
||||||||||||||||||||
Db 1 GACGTCGACCTACGCGCTGAC 21

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 21; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACGTCGACCTACGCGCTGAC 21
||||||||||||||||||||
Db 987 GACGTCGACCTACGCGCTGAC 1007

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 10 .0%; Score 21; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACGTCGACCTACGCGCTGAC 21
||||||||||||||||||||
Db 1002 GACGTCGACCTACGCGCTGAC 1022

RESULT 5
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions. AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match           100.0%;  Score 21;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 0.18;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||||||||||||
Db   1002 GACGTCGACCTACGCGCTGAC 1022


RESULT   6
Q92232/c
ID   Q92232 standard; DNA; 2538 BP.
AC   Q92232;
DT   11-JAN-1996  (first entry)
DE   Brevibacterium flavum SecA coding sequence.
KW   SecA protein; Corynform; Brevibacterium flavum; PCR; primer;
KW   amplification; probe; E.coli; membrane protein; secretory; ds.
OS   Brevibacterium flavum MJ-233.
PN   J07107981-A.
PD   25-APR-1995.
PF   08-OCT-1993; 253303.
PR   08-OCT-1993; JP-253303.
PA   (MITP ) MITSUBISHI PETROCHEMICAL CO LTD.
DR   WPI; 95-190181/25.
DR   P-PSDB; R74093.
PT   DNA fragment encoding the Coryneform bacterium secA protein -
PT   isolated from Brevibacterium flavum, used to produce transformants
PT   which secrete useful proteins
PS   Claim 4; Page 6-9; 10pp; Japanese.
CC   The nucleotide sequence of the coding sequence of the SecA protein from
CC   the Corynform bacterium Brevibacterium flavum strain MJ-233. The gene
CC   was isolated by extracting genomic DNA from B.flavum and PCR amplifying
CC   a 282 bp fragment (between bases 373-654 of this sequence), using the
CC   primers Q85959-60. The amplified fragment was used as a probe to isolate
CC   the complete gene from a B.flavum genomic library in pUC118, transformed
CC   into E.coli JM109. Transformants containing the SecA gene produce a
CC   variety of membrane proteins and secretory proteins and can be used for
CC   the production of many useful substances.
SQ   Sequence   2538 BP;    702 A;    811 C;    634 G;    391 T;


  Query Match           77.1%;  Score 16.2;  DB 1;  Length 2538;
  Best Local Similarity  85.7%;  Pred. No. 27;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          | |||||||  || ||||||||
Db   2055 GCCGTCGACGTATGCGCTGAC 2035


RESULT   7
Q10190/c
ID   Q10190 standard; DNA; 23666 BP.
AC   Q10190;
DT   27-MAR-1991  (first entry)
DE   Cephalosporin antibiotic biosynthetic genes.
KW   cephalosporin; antibiotic;
KW   S-(L-alpha-aminoadipyl)-L-cysteinyl-D-; valine synthetase;
KW   isopenicillin N synthetase; isopenicillin N epimerase;
KW   deacetoxycephalosporin C synthetase; beta-lactamase;
KW   deacetoxycephalosporin C hydroxylase; ss.
OS   Lysobacter lactamgenus.
FH   Key             Location/Qualifiers
FT   cds             5524. .16692
FT                   /*tag= a
FT                   /label= ORF 1
FT   cds             16761. .17741
FT                   /*tag= b
FT                   /label= ORF 2
FT   cds             17802. .18761
FT                   /*tag= c
FT                   /label= ORF 3
FT   cds             18798. .19739
FT                   /*tag= d
FT                   /label= ORF 4
FT   cds             19802. .21061
FT                   /*tag= e
FT                   /label= ORF 5
FT   cds             21186. .22343
FT                   /*tag= f
FT                   /label= ORF 6
FT   cds             4211. .5443
FT                   /*tag= g
FT                   /label= ORF 7
FT   cds             2887. .4200
FT                   /*tag= h
FT                   /label= ORF 8
FT   cds             6. .2819
FT                   /*tag= i
FT                   /label= ORF 9
PN   J02291274-A.
PD   03-DEC-1990.
PF   10-JAN-1990; 003762.
PR   01-FEB-1989; JP-024710.
PR   10-JAN-1990; JP-003762.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
DR   WPI; 91-018854/03.
DR   P-PSDB; R10145,R10688-R10695.
PT   Prepn. of cephalosporin series antibiotics - comprises culturing
PT   transformant of microbe transformed by plasmid contg. new DNA
PT   fragment
PS   Claim 4; Fig 1; 67pp; Japanese.
CC   A fragment of the sequence comprising at least one of the
CC   cephalosporin biosynthetic enzymes listed in the KEYWORDS can be
CC   cloned in a plasmid and used to transform microbes, such as
CC   bacteria or yeast. Although DNA is preferably isolated from L.
CC   lactamgenus, similar sequences could be obtained from other
CC   bacteria containing genes coding for biosynthesis of cephalosporin
CC   series antibiotics. See also Q10191-2.
SQ   Sequence   23666 BP;    3952 A;    8522 C;    7603 G;    3589 T;


  Query Match           77.1%;  Score 16.2;  DB 1;  Length 23666;
  Best Local Similarity  85.7%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGAC 21
          |||||||||||  |  ||||||
Db  18428 GACGTCGACCTGCAGGCTGAC 18408


RESULT   8
T93095
ID   T93095 standard; cDNA; 24379 BP.
AC   T93095;
DT   11-MAY-1998  (first entry)
DE   Streptomyces frenolicin gene cluster.
KW   Frenolicin; antibiotic; feed additive; anticoccidial;
KW   coccidiostatic; efflux pump; butyrate starter synthase;
KW   polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW   dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;
KW   ds.
OS   Streptomyces sp.
FH   Key             Location/Qualifiers
FT   CDS             636. .2948
```

```
FT                   /*tag=  a
FT                   /product= 80 kDa non-membrane protein
FT                   /note= "gene A (specifically claimed)"
FT   CDS             2945. .3916
FT                   /*tag=  b
FT                   /product= membrane protein
FT                   /note= "gene B (specifically claimed)"
FT   CDS             4020. .4844
FT                   /*tag=  c
FT                   /product= protein with 6 membrane-spanning domains
FT                   /note= "gene C (specifically claimed)"
FT   CDS             4841. .6415
FT                   /*tag=  d
FT                   /product= ATP-binding component of ABC transporter
FT                   /note= "gene D (specifically claimed)"
FT   CDS             6533. .7183
FT                   /*tag=  e
FT                   /product= unknown non-membrane protein
FT                   /note= "gene E (specifically claimed)"
FT   CDS             7344. .8897
FT                   /*tag=  f
FT                   /product= putative efflux pump
FT                   /note= "gene F (specifically claimed)"
FT   CDS             9164. .10012
FT                   /*tag=  g
FT                   /product= transcription activator
FT                   /note= "gene G (specifically claimed)"
FT   CDS             complement (10105. .10621)
FT                   /*tag=  h
FT                   /product= translationally coupled to gene I
FT                   /note= "gene H (specifically claimed)"
FT   CDS             complement (10618. .11628)
FT                   /*tag=  i
FT                   /product= homologue of fabH
FT                   /note= "gene I (specifically claimed)"
FT   CDS             11809. .12066
FT                   /*tag=  j
FT                   /product= acyl carrier protein
FT                   /note= "gene J (specifically claimed)"
FT   CDS             complement (12154. .13209)
FT                   /*tag=  k
FT                   /product= acyltransferase
FT                   /note= "gene K (specifically claimed)"
FT   CDS             13409. .14686
FT                   /*tag=  l
FT                   /product= PKS ketoacylsynthase subunit
FT                   /note= "gene L (specifically claimed)"
FT   CDS             14767. .16047
FT                   /*tag=  m
FT                   /product= PKS chain length factor
FT                   /note= "gene M (specifically claimed)"
FT   CDS             16120. .16371
FT                   /*tag=  n
FT                   /product= acyl carrier protein used by the PKS
FT                   /note= "gene N (specifically claimed)"
FT   CDS             complement (16453. .16935)
FT                   /*tag=  o
FT                   /product= putative hemiketal dehydrase
FT                   /note= "gene O (specifically claimed)"
FT   CDS             17088. .17903
FT                   /*tag=  p
FT                   /product= ketoreductase related to actIII
FT                   /note= "gene P (specifically claimed)"
FT   CDS             17903. .18898
FT                   /*tag=  q
FT                   /product= cyclase/dehydrase related to act VII
FT                   /note= "gene Q (specifically claimed)"
FT   CDS             18895. .19839
FT                   /*tag=  r
FT                   /product= cyclase/dehydrase related to actIV
FT                   /note= "gene R (specifically claimed)"
FT   CDS             complement (19990. .20907)
FT                   /*tag=  s
FT                   /product= oxidoreductase
FT                   /note= "gene S (specifically claimed)"
FT   CDS             complement (20904. .22094)
FT                   /*tag=  t
FT                   /product= quinone-forming hydroxylase
FT                   /note= "gene T (specifically claimed)"
FT   CDS             complement (22505. .22179)
FT                   /*tag=  u
FT                   /product= not specified
FT                   /note= "gene U"
PN   EP-806480-A2.
PD   12-NOV-1997.
PF   02-MAY-1997; 107329.
PR   07-MAY-1996; US-016753.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Reeves CD, Soliday CL;
DR   WPI; 97-538619/50.
DR   P-PSDB; W34199-W34219.
PT   Streptomyces frenolicin gene cluster - useful for producing
PT   recombinant frenolicin antibiotics
PS   Claim 1; Page 40-60; 66pp; English.
CC   This DNA sequence comprises the Streptomyces frenolicin gene
CC   cluster containing specifically claimed coding sequences (genes
CC   A-U) that respectively encode 21 proteins (see W34199-219) involved
CC   in frenolicin synthesis.  The genes can be divided into 5
CC   subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC   genes H, I, J and K encode butyrate starter synthases; (3) genes
CC   L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC   R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC   and (5) genes S and T encode a keto/enoyl reductase and a
CC   hydrolase.  Also claimed are vectors, host cells (especially a
CC   Streptomyces sp., particularly Streptomyces roseofulvus), and the
CC   encoded proteins.  Cells transformed using the above sequence can
CC   be cultured to produce frenolicins or frenolicin precursors.  The
CC   precursors can be converted to frenolicins by chemical or other
CC   methods.  The frenolicins can be oxidised to frenolicin B, an
CC   antibiotic used as an anticoccidial agent.  The frenolicins can be
CC   used as animal feed additives.
SQ   Sequence   24379 BP;    3077 A;    9792 C;    8499 G;    3011 T;


  Query Match             77.1%;  Score 16.2;  DB 1;  Length 24379;
  Best Local Similarity   85.7%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGAC 21
           | || ||||| ||||||||||
Db    1352 GGCGGCGACCGACGCGCTGAC 1372


RESULT   9
V25925
ID   V25925 standard; cDNA; 24379 BP.
AC   V25925;
DT   15-JUL-1998  (first entry)
DE   Streptomyces roseofulvus frenolicin gene cluster.
KW   Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
KW   antibiotic; ss.
OS   Streptomyces roseofulvus.
FH   Key             Location/Qualifiers
FT   CDS             636. .2948
FT                   /*tag=  a
FT                   /transl_except= (pos:636. .638,aa:Met)
FT                   /note= "encodes protein given in W55800"
FT   CDS             2945. .3916
FT                   /*tag=  b
FT                   /transl_except= (pos:2945. .2947,aa:Met)
FT                   /note= "encodes protein given in W55801"
FT   CDS             4020. .4844
FT                   /*tag=  c
FT                   /transl_except= (pos:4020. .4022,aa:Met)
FT                   /note= "encodes protein given in W55802"
FT   CDS             4841. .6415
```

```
FT                   /*tag= d
FT                   /note= "encodes protein given in W55803"
FT   CDS             6533. .7183
FT                   /*tag= e
FT                   /transl_except= (pos:6533. .6535,aa:Met)
FT                   /note= "encodes protein given in W55804"
FT   CDS             7344. .8897
FT                   /*tag= f
FT                   /note= "encodes protein given in W55805"
FT   CDS             9164. .10012
FT                   /*tag= g
FT                   /transl_except= (pos:9164. .9166,aa:Met)
FT                   /note= "encodes protein given in W55806"
FT   CDS             complement (10105. .10621)
FT                   /*tag= h
FT                   /transl_except= (pos:10621. .10619,aa:Met)
FT                   /note= "encodes protein given in W55807"
FT   CDS             complement (10618. .11628)
FT                   /*tag= i
FT                   /transl_except= (pos:11628. .11626,aa:Met)
FT                   /note= "encodes protein given in W55808"
FT   CDS             11809. .12066
FT                   /*tag= j
FT                   /note= "encodes protein given in W55809"
FT   CDS             complement (12154. .13209)
FT                   /*tag= k
FT                   /transl_except= (pos:13209. .13207,aa:Met
FT                   /note= "encodes protein given in W55810"
FT   CDS             13409. .14686
FT                   /*tag= l
FT                   /transl_except= (pos:13409. .13411,aa:Met)
FT                   /note= "encodes protein given in W55811"
FT   CDS             14767. .16047
FT                   /*tag= m
FT                   /note= "encodes protein given in W55812"
FT   CDS             16120. .16371
FT                   /*tag= n
FT                   /note= "encodes protein given in W55813"
FT   CDS             complement (16453. .16935)
FT                   /*tag= o
FT                   /transl_except= (pos:16935. .16933,aa:Met)
FT                   /note= "encodes protein given in W55814"
FT   CDS             17088. .18903
FT                   /*tag= p
FT                   /note= "encodes protein given in W55815"
FT   CDS             17903. .18898
FT                   /*tag= q
FT                   /note= "encodes protein given in W55816"
FT   CDS             18895. .19839
FT                   /*tag= r
FT                   /note= "encodes protein given in W55817"
FT   CDS             complement (19990. .20907)
FT                   /*tag= s
FT                   /transl_except= (pos:20907. .20905,aa:Met)
FT                   /note= "encodes protein given in W55818"
FT   CDS             complement (20904. .22094)
FT                   /*tag= t
FT                   /transl_except= (pos:22040. .22038,aa:Arg)
FT                   /transl_except= (pos:22037. .22035,aa:Arg)
FT                   /transl_except= (pos:22034. .22032,aa:Thr)
FT                   /transl_except= (pos:22031. .22029,aa:Lys)
FT                   /note= "encodes protein given in W55819"
PN   J10094395-A.
PD   14-APR-1998.
PF   07-MAY-1997; 116652.
PR   04-APR-1997; US-042935.
PR   07-MAY-1996; US-016753.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
DR   WPI; 98-279231/25.
DR   P-PSDB; W55800-W55819.
PT   Frenolicin gene cluster - useful as an antibiotic
PS   Claim 2; Page 36-44; 50pp; Japanese.
CC   The present sequence represents a frenolicin gene cluster from
CC   Streptomyces roseofulvus. The present invention describes: (1) a
CC   vector containing the frenolicin gene cluster DNA sequence operably
CC   connected to an expression control sequence; (2) a host cell
CC   transformed by the above vector; (3) a protein coded by the above DNA
CC   sequence; (4) a method for the preparation of frenolicin or a
CC   biosynthetic intermediate for it in which the above cell is cultured
CC   and frenolicin or its biosynthetic intermediate is isolated from the
CC   culture or the cell; (5) a method for the preparation of frenolicin B
CC   by oxidising frenolicin, and (6) a method for the preparation of a
CC   feed composition by mixing frenolicin with other components.
CC   Frenolicin B is useful as an antibiotic.
SQ   Sequence   24379 BP;    3077 A;    9792 C;    8499 G;    3011 T;


  Query Match             77.1%;  Score 16.2;  DB 1;  Length 24379;
  Best Local Similarity   85.7%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGAC 21
           | || ||||| ||||||||||
Db    1352 GGCGGCGACCGACGCGCTGAC 1372


RESULT  10
Q37959
ID   Q37959 standard; DNA; 358 BP.
AC   Q37959;
DT   06-JUL-1993  (first entry)
DE   Streptomyces ghanaensis phage I19 p14 promoter.
KW   Strongly constitutive promoters; lytic bacteriophage; isolation;
KW   identification; promoter search vectors; ds.
OS   Streptomyces ghanaensis phage I19.
FH   Key             Location/Qualifiers
FT   -35_signal      151. .156
FT                   /*tag= a
FT                   /note= "-35 II"
FT   -35_signal      157. .162
FT                   /*tag= b
FT                   /note= "-35 I"
FT   -10_signal      176. .181
FT                   /*tag= c
FT                   /note= "-10 II"
FT   -10_signal      180. .185
FT                   /*tag= d
FT                   /note= "-10 I"
FT   rbs             199. .207
FT                   /*tag= e
FT                   /note= "RBSI"
FT   rbs             266. .277
FT                   /*tag= f
FT                   /note= "RBS II"
PN   EP-531717-A.
PD   17-MAR-1993.
PF   07-AUG-1992; 113466.
PR   09-AUG-1991; DE-126415.
PA   (FARH ) HOECHST AG.
PI   Labes G, Wohlleben W;
DR   WPI; 93-086817/11.
PT   E. coli and Streptomyces vector for searching promoters - using
PT   promoter free gentamycin resistance gene as indicator, useful in
PT   identifying promoters strongly expressed in Streptomyces
PS   Example; Page 5; 11pp; German.
CC   The sequence is that of promoter p14 from Streptomyces ghanaensis phage
CC   I19, a lytic bacteriophage. It was isolated from this phage using a
CC   promoter search vector which uses a promoter free gentamycin gene as
CC   an indicator. By isolating strongly constitutive promoters
CC   Streptomyces can be used for the expression of any genes. Using such
CC   a vector, fragments carrying a promoter can directly be tested in
CC   E.coli for their activity without recloning.
SQ   Sequence   358 BP;    91 A;    82 C;    110 G;    75 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 358;
```

```
  Best Local Similarity   85.0%;  Pred. No. 85;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||| |||||||||  ||||
Db     136 ACGTAGACCTACGCCTTGAC 155


RESULT  11
V20280/c
ID   V20280 standard; DNA; 398 BP.
AC   V20280;
DT   26-JUN-1998  (first entry)
DE   Probe (181) for microbial genes induced during host infection.
KW   Probe; identification; microbial gene; pathogenic microorganism;
KW   host infection; virulence gene; vaccine; antimicrobial agent; ss.
OS   Salmonella typhimurium.
PN   WO9744487-A1.
PD   27-NOV-1997.
PF   16-MAY-1997; U08208.
PR   17-MAY-1996; US-651155.
PA   (REGC ) UNIV CALIFORNIA.
PI   Conner CP, Heithoff DM, Mahan MJ;
DR   WPI; 98-018538/02.
PT   Identification of microbial coding sequences - for use in vaccines
PT   against virulent pathogenic microorganisms
PS   Claim 3; Page 37; 172pp; English.
CC   The present sequence, which is derived from a Salmonella
CC   typhimurium gene specifically induced during host infection, can be
CC   used as a probe for the identification of microbial genes
CC   specifically induced in a pathogenic microorganism during host
CC   infection, i.e. virulence genes. The products of the identified
CC   virulence genes provide targets for the development of vaccines or
CC   antimicrobial agents. The vaccines can be used to protect a
CC   mammalian host against the pathogenesis of virulent microorganisms.
SQ   Sequence   398 BP;    84 A;    104 C;    112 G;    98 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 398;
  Best Local Similarity   85.0%;  Pred. No. 85;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           || ||||||||| |||| |
Db     292 ACCTCGACCTACGGGCTGCC 273


RESULT  12
V73507
ID   V73507 standard; DNA; 1023 BP.
AC   V73507;
DT   23-FEB-1999  (first entry)
DE   E. chrysanthemi HrpN-ECH DNA #2.
KW   Hypersensitive response elicitor protein; hrpN-Ech; pathogen resistance;
KW   plant; transformation; pathogen-inducible promoter; ss.
OS   Erwinia chrysanthemi.
PN   US5850015-A.
PD   15-DEC-1998.
PF   07-JUN-1995; 484358.
PR   07-JUN-1995; US-484358.
PA   (CORR ) CORNELL RES FOUND INC.
PI   Bauer D, Collmer A;
DR   WPI; 99-069852/06.
DR   P-PSDB; W82407.
PT   DNA encoding Erwinia chrysanthemi hypersensitive response protein
PT   hrpN - useful for imparting pathogen resistance to plants
PS   Claim 1; Column 31-34; 27pp; English.
CC   This sequence encodes a novel Erwinia chrysanthemi protein, hrpN-Ech,
CC   that elicits a hypersensitive response in plants. The encoding DNA can be
CC   used for imparting pathogen resistance to plants, by transforming a
CC   plant with a vector containing the DNA and a pathogen-inducible promoter.
SQ   Sequence   1023 BP;    238 A;    254 C;    331 G;    200 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1023;
  Best Local Similarity   85.0%;  Pred. No. 81;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           | || ||||| |||||||||
Db     142 AAGTTGACCTCCGCGCTGAC 161


RESULT  13
N91423
ID   N91423 standard; DNA; 1125 BP.
AC   N91423;
DT   01-FEB-1991  (first entry)
DE   Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt.
DE   2 MD or less
KW   Enzyme; ds.
OS   Bacillus stearothermophilus IFO 12550.
PN   J01043194-A.
PD   15-FEB-1989.
PF   10-AUG-1987; 200524.
PR   10-AUG-1987; JP-200524.
PA   (NIRA) Unitika Kk.
DR   WPI; 89-096096/13.
PT   Recombinant plasmid for transforming Escherichia coli -
PT   obtd. by connecting heat resistant alanine dehydrogenase gene to
PT   vector plasmid
PS   Disclosure; Fig 2, p693 ; 9pp; Japanese.
CC   A recombinant plasmid contg. heat-resistant alanine dehydrogenase (AH)
CC   gene with mol. wt. 2 or less MD is claimed.  Cells transformed with the
CC   vector produce high levels of heat-resistant AH.  Transformed E. coli
CC   is useful as a clinical inspection reagent.
SQ   Sequence   1125 BP;    246 A;    295 C;    367 G;    217 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1125;
  Best Local Similarity   85.0%;  Pred. No. 81;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||||| | ||||||  ||
Db     922 ACGTCGACATTCGCGCTTAC 941


RESULT  14
T92239
ID   T92239 standard; DNA; 1746 BP.
AC   T92239;
DT   19-JAN-1998  (first entry)
DE   Mercuric ion reductase gene merApe29.
KW   Metal ion resistance; mercuric ion reductase; merA; mercury;
KW   toxic heavy metal; transgenic plant; mutagenic; bacterial;
KW   transposon; Tn21; ds.
OS   Transposon Tn21.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   CDS             40. .1728
FT                   /*tag= a
FT                   /product= Mutagenised_merApe29
FT   mat_peptide     40. .1725
FT                   /*tag= b
PN   US5668294-A.
PD   16-SEP-1997.
PF   21-APR-1995; 427097.
PR   21-APR-1995; US-427097.
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
PI   Meagher RB,  Summers AO;
DR   WPI; 97-470112/43.
DR   P-PSDB; W32515.
PT   DNA encoding mercuric ion reductase in plant-expressible form - for
PT   producing transgenic plants resistant to toxic heavy metals
PS   Claim 1; Column 81-86; 52pp; English.
```

CC The present sequence represents a nucleic acid molecule comprising a
CC sequence encoding a metal ion resistance gene, where the gene has the
CC coding sequence merApe29. The nucleic acid molecule is useful for
CC producing transgenic plants that are resistant to toxic heavy metals
CC (especially mercury) and so can be used remediate and/or revegetate
CC contaminated areas. The merApe29 gene is derived by mutating the
CC bacterial merA gene which codes for mercuric ion reductase. The
CC bacterial merA gene is derived from the transposon Tn21, which was
CC originally isolated from the Incompatibility Group IncFII resistance
CC plasmid NR1.
SQ Sequence 1746 BP; 363 A; 549 C; 520 G; 314 T;

Query Match 72.4%; Score 15.2; DB 1; Length 1746;
Best Local Similarity 85.0%; Pred. No. 79;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ACGTCGACCTACGCGCTGAC 21
||||| || ||||||||||
Db 181 ACGTCACCCGACGCGCTGAC 200

RESULT 15
T92237
ID T92237 standard; DNA; 1752 BP.
AC T92237;
DT 19-JAN-1998 (first entry)
DE Mercuric ion reductase gene merApe9.
KW Metal ion resistance; mercuric ion reductase; merA; mercury;
KW toxic heavy metal; transgenic plant; mutagenic; bacterial;
KW transposon; Tn21; ds.
OS Transposon Tn21.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 40. .1734
FT /*tag= a
FT /product= Mutagenised_merApe9
FT mat_peptide 40. .1731
FT /*tag= b
PN US5668294-A.
PD 16-SEP-1997.
PF 21-APR-1995; 427097.
PR 21-APR-1995; US-427097.
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PI Meagher RB, Summers AO;
DR WPI; 97-470112/43.
DR P-PSDB; W32513.
PT DNA encoding mercuric ion reductase in plant-expressible form - for
PT producing transgenic plants resistant to toxic heavy metals
PS Claim 1; Column 55-60; 52pp; English.
CC The present sequence represents a nucleic acid molecule comprising a
CC sequence encoding a metal ion resistance gene, where the gene has the
CC coding sequence merApe9. The nucleic acid molecule is useful for
CC producing transgenic plants that are resistant to toxic heavy metals
CC (especially mercury) and so can be used remediate and/or revegetate
CC contaminated areas. The merApe9 gene is derived by mutating the
CC bacterial merA gene which codes for mercuric ion reductase. The
CC bacterial merA gene is derived from the transposon Tn21, which was
CC originally isolated from the Incompatibility Group IncFII resistance
CC plasmid NR1.
SQ Sequence 1752 BP; 356 A; 564 C; 540 G; 292 T;

Query Match 72.4%; Score 15.2; DB 1; Length 1752;
Best Local Similarity 85.0%; Pred. No. 79;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ACGTCGACCTACGCGCTGAC 21
||||| || ||||||||||
Db 181 ACGTCACCCGACGCGCTGAC 200

Search completed: June 22, 2000, 15:12:54
Job time: 18100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22,. 2000, 15:22:00 ; Search time 341.15 Seconds
(without alignments)
8.001 Million cell updates/sec

Title: US-09-362-485-17
Perfect score: 21
Sequence: 1 GACGTCGACCTACGCGCTGAC 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	16.8	80.0	3606	1	US-07-661-610C-7	Sequence 7, Appli
c	2	15.2	72.4	33	4	US-08-324-003A-14	Sequence 14, Appl
	3	15.2	72.4	358	1	US-07-925-920-1	Sequence 1, Appli
	4	15.2	72.4	1023	3	US-08-484-358-6	Sequence 6, Appli
	5	15.2	72.4	1023	5	US-09-118-959-6	Sequence 6, Appli
	6	15.2	72.4	1728	1	US-08-427-097-1	Sequence 1, Appli
	7	15.2	72.4	1728	4	US-08-878-957-1	Sequence 1, Appli
	8	15.2	72.4	1746	1	US-08-427-097-29	Sequence 29, Appl
	9	15.2	72.4	1746	4	US-08-878-957-29	Sequence 29, Appl
	10	15.2	72.4	1752	1	US-08-427-097-13	Sequence 13, Appl
	11	15.2	72.4	1752	1	US-08-427-097-15	Sequence 15, Appl
	12	15.2	72.4	1752	1	US-08-427-097-19	Sequence 19, Appl
	13	15.2	72.4	1752	1	US-08-427-097-27	Sequence 27, Appl
	14	15.2	72.4	1752	4	US-08-878-957-13	Sequence 13, Appl
	15	15.2	72.4	1752	4	US-08-878-957-15	Sequence 15, Appl
	16	15.2	72.4	1752	4	US-08-878-957-19	Sequence 19, Appl
	17	15.2	72.4	1752	4	US-08-878-957-27	Sequence 27, Appl
	18	15.2	72.4	2109	3	US-08-617-801A-5	Sequence 5, Appli
	19	15.2	72.4	2141	2	US-08-891-254-2	Sequence 2, Appli
	20	15.2	72.4	2141	3	US-08-484-358-1	Sequence 1, Appli
	21	15.2	72.4	2141	3	US-08-819-539-2	Sequence 2, Appli
	22	15.2	72.4	2141	4	US-09-030-270A-2	Sequence 2, Appli
	23	15.2	72.4	2141	5	US-09-118-959-1	Sequence 1, Appli
	24	15.2	72.4	2191	6	PCT-US96-08819-2	Sequence 2, Appli
	25	15.2	72.4	2220	3	US-08-617-801A-1	Sequence 1, Appli
	26	15.2	72.4	2245	3	US-08-617-801A-3	Sequence 3, Appli
c	27	14.6	69.5	1254	1	US-08-313-553-1	Sequence 1, Appli
c	28	14.6	69.5	1254	5	US-08-767-993-1	Sequence 1, Appli
	29	14.6	69.5	1268	3	US-08-403-852D-2	Sequence 2, Appli
c	30	14.6	69.5	1632	1	US-08-362-232-1	Sequence 1, Appli
c	31	14.6	69.5	1632	2	US-08-814-196-1	Sequence 1, Appli
c	32	14.6	69.5	2147	1	US-08-313-553-14	Sequence 14, Appl
c	33	14.6	69.5	2147	5	US-08-767-993-14	Sequence 14, Appl
	34	14.6	69.5	5392	3	US-08-403-852D-1	Sequence 1, Appli
c	35	14.4	68.6	1121	4	US-08-841-349-13	Sequence 13, Appl
	36	14.2	67.6	267	6	PCT-US96-04648-3	Sequence 3, Appli
	37	14.2	67.6	441	1	US-08-217-918-3	Sequence 3, Appli
c	38	14.2	67.6	618	2	US-08-800-751-3	Sequence 3, Appli
c	39	14.2	67.6	618	4	US-08-990-818-3	Sequence 3, Appli
c	40	14.2	67.6	5924	2	US-08-459-586-6	Sequence 6, Appli
c	41	14.2	67.6	5924	3	US-08-282-696-6	Sequence 6, Appli
	42	14.2	67.6	10718	5	US-08-325-426B-1	Sequence 1, Appli
c	43	14.2	67.6	49377	2	US-08-764-233A-1	Sequence 1, Appli
	44	14	66.7	30001	1	US-08-125-468-1	Sequence 1, Appli
	45	14	66.7	30001	3	US-08-474-933-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-07-661-610C-7/c
; Sequence 7, Application US/07661610C
; Patent No. 5292643
;  GENERAL INFORMATION:
;    APPLICANT:  Shibano, Yuji
;    APPLICANT:  Toyoda, Hideyoshi
;    APPLICANT:  Utsumi, Ryutaro
;    APPLICANT:  Obata, Kazuaki
;    TITLE OF INVENTION:  Fusaric Acid Resistant Genes
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/661,610C
;      FILING DATE:  19910228
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5292643man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2292-010-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3606 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  385..813
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1382..2083
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  2591..3041
US-07-661-610C-7


  Query Match             80.0%;  Score 16.8;  DB 1;  Length 3606;
  Best Local Similarity   90.0%;  Pred. No. 7.7;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGA 20
           ||||||| ||| ||||||||
Db    2803 GACGTCGGCCTGCGCGCTGA 2784


RESULT   2
US-08-324-003A-14/c
; Sequence 14, Application US/08324003A
; Patent No. 5977438
;  GENERAL INFORMATION:
;    APPLICANT:  Turpen, Thomas H.
;    APPLICANT:  Reinl, Stephen
;    APPLICANT:  Grill, Laurence K.
;    TITLE OF INVENTION:  Production of Peptides in Plants as
;    TITLE OF INVENTION:  Viral Coat Protein Fusions
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/324,003A
;      FILING DATE:  14-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Halluin, Albert P.
;      REGISTRATION NUMBER:  25,227
;      REFERENCE/DOCKET NUMBER:  8129-087
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-3660
;      TELEFAX:  415-854-3694
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  33 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-324-003A-14


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 33;
  Best Local Similarity   85.0%;  Pred. No. 50;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||||||| | | ||||
Db      33 ACGTCGACCTAGGTGATGAC 14


RESULT   3
US-07-925-920-1
; Sequence 1, Application US/07925920
; Patent No. 5328998
;  GENERAL INFORMATION:
;    APPLICANT:  Labes, Gabrielle
;    APPLICANT:  Wohlleben, Wolfgang
;    TITLE OF INVENTION:  A Promoter Screening Vector,
;    TITLE OF INVENTION:  Streptomyces Promoters Found Therewith, and the Isolation
;    TITLE OF INVENTION:  and Use Thereof
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/925,920
;      FILING DATE:  19920807
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 41 26 415.0
;      FILING DATE:  09-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wadler, Linda A.
;      REGISTRATION NUMBER:  33,218
;      REFERENCE/DOCKET NUMBER:  02481-1213-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  358 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-925-920-1


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 358;
  Best Local Similarity   85.0%;  Pred. No. 45;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||| |||||||||  ||||
Db     136 ACGTAGACCTACGCCTTGAC 155


RESULT   4
US-08-484-358-6
; Sequence 6, Application US/08484358
; Patent No. 5850015
;  GENERAL INFORMATION:
;    APPLICANT:  Bauer, David
;    APPLICANT:  Collmer, Alan
;    TITLE OF INVENTION:  Hypersensitive Response Elicitor
;    TITLE OF INVENTION:  From
;    TITLE OF INVENTION:  Erwinia Chrysanthemi
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,358
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/840
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1023 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-484-358-6


  Query Match             72.4%;  Score 15.2;  DB 3;  Length 1023;
  Best Local Similarity   85.0%;  Pred. No. 44;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           | || ||||| |||||||||
Db     142 AAGTTGACCTCCGCGCTGAC 161


RESULT   5
US-09-118-959-6
; Sequence 6, Application US/09118959
; Patent No. 6001959
;  GENERAL INFORMATION:
;    APPLICANT:  Bauer, David
;    APPLICANT:  Collmer, Alan
;    TITLE OF INVENTION:  Hypersensitive Response Elicitor From
;    TITLE OF INVENTION:  Erwinia Chrysanthemi
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/118,959
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/840
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1023 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-118-959-6


  Query Match             72.4%;  Score 15.2;  DB 5;  Length 1023;
  Best Local Similarity   85.0%;  Pred. No. 44;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           | || ||||| |||||||||
Db     142 AAGTTGACCTCCGCGCTGAC 161


RESULT   6
US-08-427-097-1
; Sequence 1, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1728 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1708
US-08-427-097-1


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1728;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     155 ACGTCACCCGACGCGCTGAC 174


RESULT   7
US-08-878-957-1
; Sequence 1, Application US/08878957
; Patent No. 5965796
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
```

```
;    APPLICANT:  Summers, Anne O.
;    APPLICANT:  Rugh, Clayton L.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,957
;      FILING DATE:  19-JUN-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/427,097
;      FILING DATE:  21-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1728 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1708
US-08-878-957-1


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 1728;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     155 ACGTCACCCGACGCGCTGAC 174


RESULT   8
US-08-427-097-29
; Sequence 29, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe29"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1728
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1725
US-08-427-097-29


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1746;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT   9
US-08-878-957-29
; Sequence 29, Application US/08878957
; Patent No. 5965796
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Summers, Anne O.
;    APPLICANT:  Rugh, Clayton L.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,957
;      FILING DATE:  19-JUN-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/427,097
;      FILING DATE:  21-APR-1995
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe29"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1728
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1725
US-08-878-957-29


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 1746;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  10
US-08-427-097-13
; Sequence 13, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe38"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-427-097-13


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  11
US-08-427-097-15
; Sequence 15, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe9"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-427-097-15
```

```
  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  12
US-08-427-097-19
; Sequence 19, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe47"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-427-097-19


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  13
US-08-427-097-27
; Sequence 27, Application US/08427097
; Patent No. 5668294
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Sommers, Anne O.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/427,097
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe20"
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-427-097-27


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  14
US-08-878-957-13
; Sequence 13, Application US/08878957
; Patent No. 5965796
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Summers, Anne O.
;    APPLICANT:  Rugh, Clayton L.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
```

```
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,957
;      FILING DATE:  19-JUN-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/427,097
;      FILING DATE:  21-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe38"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-878-957-13


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || ||||||||||
Db     181 ACGTCACCCGACGCGCTGAC 200


RESULT  15
US-08-878-957-15
; Sequence 15, Application US/08878957
; Patent No. 5965796
;  GENERAL INFORMATION:
;    APPLICANT:  Meagher, Richard B.
;    APPLICANT:  Summers, Anne O.
;    APPLICANT:  Rugh, Clayton L.
;    TITLE OF INVENTION:  Metal Resistance Sequences and
;    TITLE OF INVENTION:  Transgenic Plants
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,957
;      FILING DATE:  19-JUN-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/427,097
;      FILING DATE:  21-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  40-94A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Mutagenized merApe9"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  40..1734
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  40..1731
US-08-878-957-15


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 1752;
  Best Local Similarity   85.0%;  Pred. No. 43;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           |||||  || |||||||||
Db     181 ACGTCACCCGACGCGCTGA  200


Search completed: June 22, 2000, 15:22:01
Job time: 18799 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:30 ; Search time 5541.94 Seconds
(without alignments)
15.359 Million cell updates/sec

Title: US-09-362-485-17
Perfect score: 21
Sequence: 1 GACGTCGACCTACGCGCTGAC 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query %

	No.	Score	Match	Length	DB	ID	Description
	1	16.2	77.1	619	82	FR0019828	AL012715 F.rubripe
	2	15.8	75.2	267	25	W05201	W05201 za42g08.r1
	3	15.8	75.2	335	91	AQ853324	AQ853324 LMAJFV1_1
	4	15.8	75.2	347	21	T62468	T62468 tgEST0232 T
	5	15.8	75.2	537	34	AA531750	AA531750 TgESTzz31
	6	15.4	73.3	518	103	AQ250254	AQ250254 T1B18-Sp6
	7	15.2	72.4	134	20	T24793	T24793 EST368 Huma
c	8	15.2	72.4	240	35	C42852	C42852 C42852 Yuji
	9	15.2	72.4	264	34	C25439	C25439 C25439 Rice
c	10	15.2	72.4	360	35	C40405	C40405 C40405 Yuji
c	11	15.2	72.4	360	35	C47471	C47471 C47471 Yuji
	12	15.2	72.4	364	60	AI808524	AI808524 wf95e04.x
c	13	15.2	72.4	443	74	AW199997	AW199997 da10d02.y
c	14	15.2	72.4	449	63	AI948101	AI948101 603034F05
	15	15.2	72.4	494	41	AI066550	AI066550 oo14f11.x
c	16	15.2	72.4	501	50	AI673923	AI673923 605039B09
	17	15.2	72.4	517	85	AQ681804	AQ681804 HS_5502_B
	18	15.2	72.4	523	103	AQ156539	AQ156539 nbxb0008E
	19	15.2	72.4	569	91	AQ845940	AQ845940 LMAJFV1_1
c	20	15.2	72.4	578	48	AI621681	AI621681 486006C10
	21	15.2	72.4	595	44	AI297302	AI297302 LP11571.5
c	22	15.2	72.4	623	63	AI980321	AI980321 pat.pk002
	23	15.2	72.4	650	63	AI943517	AI943517 BSBmMFSZ4
	24	15.2	72.4	655	74	AW179565	AW179565 SWYD25CAU
	25	15.2	72.4	708	103	AQ256391	AQ256391 nbxb0016A
	26	15.2	72.4	989	82	CNS00L0U	AL067826 Drosophil
	27	15.2	72.4	1013	82	CNS00IEQ	AL074889 Drosophil
	28	15.2	72.4	1100	82	CNS001YP	AL061531 Drosophil
	29	15.2	72.4	1101	82	CNS00BJG	AL057254 Drosophil
c	30	15	71.4	348	46	AI028998	AI028998 UI-R-C0-i
c	31	15	71.4	396	43	AI233089	AI233089 EST229777
c	32	15	71.4	486	39	AA850543	AA850543 EST193310
	33	15	71.4	781	35	AA541066	AA541066 LD20853.5
	34	14.8	70.5	163	33	AA444091	AA444091 zv46d06.r
c	35	14.8	70.5	251	30	AA229192	AA229192 nc46h10.r
	36	14.8	70.5	358	50	AU064544	AU064544 AU064544
c	37	14.8	70.5	373	30	AA225422	AA225422 nc24a11.r
c	38	14.8	70.5	401	92	AQ907812	AQ907812 GSSTc0618
c	39	14.8	70.5	412	29	AA154700	AA154700 mr96d06.r
c	40	14.8	70.5	469	107	AQ502072	AQ502072 V9A8 mTn-
c	41	14.8	70.5	501	64	AW100221	AW100221 sd21c01.y
	42	14.8	70.5	525	74	AV395068	AV395068 AV395068
c	43	14.8	70.5	601	49	C96383	C96383 C96383 Marc
	44	14.8	70.5	629	81	AW394406	AW394406 sh31h06.y
c	45	14.8	70.5	700	92	AQ949550	AQ949550 Sheared D

ALIGNMENTS

RESULT 1
FR0019828
LOCUS FR0019828 619 bp DNA GSS 10-DEC-1997
DEFINITION F.rubripes GSS sequence, clone 036E19cH6, genomic survey sequence.
ACCESSION AL012715
VERSION AL012715.1 GI:2679083
KEYWORDS GSS; genome survey sequence.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE 1 (bases 1 to 619)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .619
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 036E19"
/clone="036E19cH6"
BASE COUNT 173 a 137 c 141 g 123 t 45 others
ORIGIN

Query Match 77.1%; Score 16.2; DB 82; Length 619;
Best Local Similarity 85.7%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GACGTCGACCTACGCGCTGAC 21
||||| ||||||||||| ||
Db 471 GACGTGGACCTACGCGCCCAC 491

RESULT 2
W05201
LOCUS W05201 267 bp mRNA EST 23-APR-1996
DEFINITION za42g08.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:295262 5' similar to gb:V00493_rna1 HEMOGLOBIN ALPHA CHAIN (HUMAN);, mRNA sequence.
ACCESSION W05201
VERSION W05201.1 GI:1277933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 267)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 8, 1995 this sequence version replaced gi:800173.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 172.
FEATURES Location/Qualifiers
source 1. .267
/organism="Homo sapiens"
/db_xref="GDB:1240183"
/db_xref="taxon:9606"
/clone="IMAGE:295262"
/clone_lib="Soares fetal liver spleen 1NFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

```
BASE COUNT       60 a     79 c     79 g     41 t      8 others
ORIGIN


  Query Match             75.2%;  Score 15.8;  DB 25;  Length 267;
  Best Local Similarity   81.0%;  Pred. No. 8e+02;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGAC 21
           || || | || ||||||||||
Db     205 GAAGTNGNCCGACGCGCTGAC 225


RESULT   3
AQ853324
LOCUS       AQ853324       335 bp    DNA             GSS       29-OCT-1999
DEFINITION  LMAJFV1_lm83c09.y1 Leishmania major FV1 random genomic library
            Leishmania major genomic clone LMAJFV1_lm83c09 5', genomic survey
            sequence.
ACCESSION   AQ853324
VERSION     AQ853324.1  GI:6141515
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Akopyants,N.S., Beverley,S.M., Clifton,S., Marra,M., Hillier,L.,
            Chinwalla,A., Pape,D., Martin,J., Wylie,T., Blistain,A.,
            Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko,I.,
            Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
            Harvey,N., McCann,R., Tsagareishvili,R., Williams,T., Jackson,Y.,
            Bowers,Y., Swaller,T., Waterston,R. and Wilson,R.
  TITLE     WashU Leishmania Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Akopyants, NS / Beverley, SM
            WashU Leishmania Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library construction: Natalia S. Akopyants, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             If using this information please cite:
             N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
            Friedlin strain V1 genome by shotgun sequencing' and the Washington
            University Genome Sequencing Center For information on obtaining
            clone material please contact: Natalia S. Akopyants Ph.D.
            (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
            (beverley@borcim.wustl.edu)
            Seq primer: -40RP from Gibco
            Class: shotgun
            High quality sequence stop: 327.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Leishmania major"
                     /strain="Friedlin strain V1"
                     /db_xref="taxon:5664"
                     /clone="LMAJFV1_lm83c09"
                     /clone_lib="Leishmania major FV1 random genomic library"
                     /lab_host="TOP10 (Invitrogen)"
                     /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
                     Genomic DNA was isolated from stationary phase cells. For
                     this library, DNA was sheared to give a tight size
                     distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
                     polymerase, dephosphorylated with Shrimp Alkaline
                     Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT       42 a     95 c    121 g     77 t
ORIGIN


  Query Match             75.2%;  Score 15.8;  DB 91;  Length 335;
  Best Local Similarity   89.5%;  Pred. No. 8.4e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CGTCGACCTACGCGCTGAC 21
           | ||||| ||||||||||
Db     244 CATCGACGTACGCGCTGAC 262


RESULT   4
T62468
LOCUS       T62468         347 bp    mRNA            EST       27-AUG-1998
DEFINITION  tgEST0232 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone tgd065
            5' end, mRNA sequence.
ACCESSION   T62468
VERSION     T62468.1  GI:665910
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Wan,K.-L., Blackwell,J.M. and Ajioka,J.W.
  TITLE     Toxoplasma gondii expressed sequence tags: insight into tachyzoite
            gene expression
  JOURNAL   Mol. Biochem. Parasitol. 75 (2), 179-186 (1995)
  MEDLINE   96258558
COMMENT     Contact: Ajioka JW
            Laboratory for Parasite Genome Analysis
            Cambridge University
            Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, UK.
            Tel: 01223333923
            Fax: 01223333923
            Email: jajioka@hgmp.mrc.ac.uk
            Clone information and library can be obtained from Jim Ajioka
            jwa@mole.bio.cam.ac.uk
            Seq primer: SK
            High quality sequence stop: 331.
FEATURES             Location/Qualifiers
     source          1. .347
                     /organism="Toxoplasma gondii"
                     /strain="RH"
                     /db_xref="taxon:5811"
                     /clone="tgd065"
                     /clone_lib="TgRH Tachyzoite cDNA"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
                     Toxoplasma RH strain tachyzoites were grown in human
                     foreskin fibroblast cultures in vitro. The library was
                     constructed by K.L. Wan, Cambridge University.  cDNAs were
                     synthesized from polyA RNAs by oligo d(T) priming and
                     directionally cloned into the EcoRI to XhoI sites of the
                     Lambda ZapII vector using the ZAP-cDNA synthesis kit
                     (Stratagene). WARNING: the library contains a small
                     percentage of cDNAs derived from the human host cells."
BASE COUNT       76 a     96 c    102 g     73 t
ORIGIN


  Query Match             75.2%;  Score 15.8;  DB 21;  Length 347;
  Best Local Similarity   89.5%;  Pred. No. 8.4e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTG 19
           | |||||||||| |||||
Db      78 GCCGTCGACCTACACGCTG 96


RESULT   5
AA531750
LOCUS       AA531750       537 bp    mRNA            EST       22-JUL-1997
DEFINITION  TgESTzz31a08.r1 TgME49 invivo Bradyzoite cDNA size selected
            Toxoplasma gondii cDNA clone tgzz31a08.r1 5' similar to TR:G495670
            G495670 SURFACE ANTIGEN 43. ;, mRNA sequence.
```

```
ACCESSION   AA531750
VERSION     AA531750.1  GI:2274456
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Hehl,A., Manger,I., Marra,M., Parmley,S., Sibley,L.D., Hillier,L.,
            Allen,M., Bowles,L., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M.,
            Le,N., Jost,S., Martin,J., Moore,B., Schellenberg,K., Steptoe,M.,
            Tan,F., Theising,B., Bowers,Y., Wylie,T., Ajioka,J.A., Aslett,M.A.,
            Wan,K.L., Wilson,R., Waterston,R. and Boothroyd J.C.
  TITLE     WashU-Stanford-PAMF-NIH Toxoplasma EST project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407270.
            Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            Contact Steve Parmley, PAMF (76424.16@compuserve.com) for
            information on clone and library availability.
            High quality sequence stop: 314.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Toxoplasma gondii"
                     /strain="ME49"
                     /db_xref="taxon:5811"
                     /clone="tgzz31a08.r1"
                     /clone_lib="TgME49 invivo Bradyzoite cDNA size selected"
                     /dev_stage="Bradyzoite"
                     /lab_host="DH10"
                     /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
                     NotI; Mature bradyzoites were obtained from infected mouse
                     brains by percoll density centrifugation. The original
                     library was constructed by Steve Parmley, Palo Alto
                     Medical Foundation. cDNAs were synthesized by priming with
                     oligo d(T) and directionally cloned into the EcoRI/NotI
                     sites of lambda gt11. Warning: the library contains a
                     small percentage of host cDNAs derived from mouse cells.
                     Inserts from this cDNA library were excised with Not1 and
                     EcoR1, size selected in a range of 0.7 - 2.0 kb and
                     subcloned into Bluescript II SK- (Adrian Hehl, Ian Manger
                     and John Boothroyd, Stanford University)"
BASE COUNT      123 a    152 c    151 g    111 t
ORIGIN


  Query Match           75.2%;  Score 15.8;  DB 34;  Length 537;
  Best Local Similarity 89.5%;  Pred. No. 9.2e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTG 19
          | ||||||||||| |||||
Db    300 GCCGTCGACCTACACGCTG 318


RESULT   6
AQ250254
LOCUS       AQ250254      518 bp    DNA             GSS       07-OCT-1998
DEFINITION  T1B18-Sp6 TAMU Arabidopsis thaliana genomic clone T1B18, genomic
            survey sequence.
ACCESSION   AQ250254
VERSION     AQ250254.1  GI:3703368
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T1B18"
                     /clone_lib="TAMU"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
BASE COUNT      147 a    121 c     73 g    173 t      4 others
ORIGIN


  Query Match           73.3%;  Score 15.4;  DB 103;  Length 518;
  Best Local Similarity 94.1%;  Pred. No. 1.4e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 ACGTCGACCTACGCGCT 18
          |||||||||||| |||||
Db    437 ACGTCGACCTATGCGCT 453


RESULT   7
T24793
LOCUS       T24793        134 bp    mRNA            EST       24-AUG-1995
DEFINITION  EST368 Human colorectal cancer Homo sapiens cDNA clone 13G10, mRNA
            sequence.
ACCESSION   T24793
VERSION     T24793.1  GI:534418
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134)
  AUTHORS   Frigerio,J.-M., Berthezene,P., Garrido,P., Ortiz,E., Barthellemy,S,
            Vasseur,S., Sastre,B., Seleznieff,I., Dagorn,J.-C. and
            Iovanna,J.-L.
  TITLE     Analysis of 2166 clones from a human colorectal cancer cDNA library
            by partial sequencing
  JOURNAL   Hum. Mol. Genet. 4, 37-43 (1995)
  MEDLINE   95227175
COMMENT     Contact: Iovanna JL
            U.315 INSERM
            46 Bd de la Gaye, F-13009 Marseille, France.
            Tel: (33) 91 82 03 15
            Fax: (33) 91 26 62 19
            Email: dagorn@arthur.citi2.fr
            This sequence is one of a series obtained by systematic sequencing
            of a colorectal cancer cDNA library.
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .134
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="13G10"
```

```
                     /clone_lib="Human colorectal cancer"
                     /lab_host="E. coli NM522"
                     /note="Vector: pT7T3D; Site_1: Eco RI; Site_2: Not I; mRNA
                     was purified from a colorectal tumour of an adult male.
                     cDNA was constructed and cloned into the pT7T3D phagemid
                     following the manufacturer intructions (Pharmacia)."
BASE COUNT       20 a     45 c     45 g     22 t      2 others
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 20;  Length 134;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+03;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGTCGACCTACGCGCTGAC 21
           ||||||||  ||| ||||||
Db      10 ACGTCGACACACGTGCTGAC 29


RESULT   8
C42852/c
LOCUS       C42852         240 bp    mRNA            EST       18-OCT-1999
DEFINITION  C42852 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk325h3 5', mRNA sequence.
ACCESSION   C42852
VERSION     C42852.1  GI:2379089
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk325h3"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       71 a     63 c     52 g     47 t      7 others
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 35;  Length 240;
  Best Local Similarity   85.0%;  Pred. No. 1.4e+03;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGA 20
           ||||| || ||||| |||||
Db     176 GACGTGGAACTACGTGCTGA 157


RESULT   9
C25439
LOCUS       C25439         264 bp    mRNA            EST       25-JUL-1997
DEFINITION  C25439 Rice shoot Oryza sativa cDNA clone S6414_2A, mRNA sequence.
ACCESSION   C25439
VERSION     C25439.1  GI:2280944
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from etiolated shoot
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692397.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="S6414_2A"
                     /clone_lib="Rice shoot"
                     /note="Etiolated shoot (8 days old)"
BASE COUNT       54 a    107 c     66 g     31 t      6 others
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 34;  Length 264;
  Best Local Similarity   81.0%;  Pred. No. 1.5e+03;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 GACGTCGACCTACGCGCTGAC 21
           | |||||||||||| | | |||
Db     100 GNCGTCGACCTACACACCGAC 120


RESULT  10
C40405/c
LOCUS       C40405         360 bp    mRNA            EST       18-OCT-1999
DEFINITION  C40405 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk230e1 5', mRNA sequence.
ACCESSION   C40405
VERSION     C40405.1  GI:2376642
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394191.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk230e1"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
```

```
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT      111 a     89 c     84 g     76 t
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 35;  Length 360;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+03;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGA 20
          ||||| || ||||| |||||
Db    206 GACGTGGAACTACGTGCTGA 187


RESULT  11
C47471/c
LOCUS       C47471         360 bp    mRNA            EST       18-OCT-1999
DEFINITION  C47471 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk447a4 5', mRNA sequence.
ACCESSION   C47471
VERSION     C47471.1  GI:2383724
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1282340.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk447a4"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT      109 a     88 c     79 g     75 t      9 others
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 35;  Length 360;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+03;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGA 20
          ||||| || ||||| |||||
Db    179 GACGTGGAACTACGTGCTGA 160


RESULT  12
AI808524
LOCUS       AI808524       364 bp    mRNA            EST       19-DEC-1999
DEFINITION  wf95e04.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:2363358 3', mRNA sequence.
ACCESSION   AI808524
VERSION     AI808524.1  GI:5395090
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189441.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 774   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2363358"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT       49 a    125 c    131 g     59 t
ORIGIN


  Query Match             72.4%;  Score 15.2;  DB 60;  Length 364;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+03;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GACGTCGACCTACGCGCTGA 20
          ||||| ||||  ||||||||
Db    229 GACGTTGACCGCCGCGCTGA 248


RESULT  13
AW199997/c
LOCUS       AW199997       443 bp    mRNA            EST       30-NOV-1999
DEFINITION  dal0d02.y1 Xenopus laevis oocyte Xenopus laevis cDNA clone
            XENOPUS_SOURCE_ID:xlnoc003f04 5', mRNA sequence.
ACCESSION   AW199997
VERSION     AW199997.1  GI:6480566
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Johnson,S.L., Blumberg,B., Song,J., Clifton,S., Hillier,L.,
            Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y.,
            Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135640.
```

Other_ESTs: da10d02.x1
Contact: Stephen L. Johnson/WashU Xenopus EST project, 1999
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for this library can be found through Research Genetics, visit their web page at: http://www.resgen.com/
Seq primer: -40RP from Gibco
High quality sequence stop: 424.

FEATURES Location/Qualifiers
source 1. .443
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS_SOURCE_ID:xlnoc003f04"
/clone_lib="Xenopus laevis oocyte"
/tissue_type="oocyte (stages 5 and 6)"
/lab_host="Top-10 F'"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. ss-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library contruction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Normalized by Jihwan Song (Song, Cho and Blumberg, unpublished)."

BASE COUNT 152 a 90 c 117 g 84 t
ORIGIN

Query Match 72.4%; Score 15.2; DB 74; Length 443;
Best Local Similarity 85.0%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 GACGTCGACCTACGCGCTGA 20
          |||||| || ||| ||||||
Db     20 GACGTCTACATACTCGCTGA 1
```

RESULT 14
AI948101/c
LOCUS AI948101 449 bp mRNA EST 19-AUG-1999
DEFINITION 603034F05.x1 603 - stressed root cDNA library from Wang/Bohnert lab Zea mays cDNA, mRNA sequence.
ACCESSION AI948101
VERSION AI948101.1 GI:5740411
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE 1 (bases 1 to 449)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT On Feb 18, 1999 this sequence version replaced gi:4296972.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603034 row: F column: 05.

FEATURES Location/Qualifiers
source 1. .449
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="603 - stressed root cDNA library from Wang/Bohnert lab"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/note="Organ: root; Vector: pBluescriptII SK(+) XR; Seedling stressed root cDNA library from Wang/Bohnert lab"

BASE COUNT 113 a 130 c 115 g 91 t
ORIGIN

Query Match 72.4%; Score 15.2; DB 63; Length 449;
Best Local Similarity 85.0%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 GACGTCGACCTACGCGCTGA 20
          |||||||||| ||||| ||
Db    375 GACGTCGACCAGCGCGCCGA 356
```

RESULT 15
AI066550
LOCUS AI066550 494 bp mRNA EST 13-FEB-1999
DEFINITION oo14f11.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:1566189 3' similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA sequence.
ACCESSION AI066550
VERSION AI066550.1 GI:3367252
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407355.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 784 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 465.

FEATURES Location/Qualifiers
source 1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1566189"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following

libraries and cloneIDs: Soares NbHSF pool 1: 309384-310919, 323208-325895 Soares Nb2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares Nb2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NbHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NbHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 65 a 171 c 179 g 79 t

ORIGIN

Query Match 72.4%; Score 15.2; DB 41; Length 494;
Best Local Similarity 85.0%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 GACGTCGACCTACGCGCTGA 20
           ||||| ||||  ||||||||
Db     256 GACGTTGACCGCCGCGCTGA 275
```

Search completed: June 22, 2000, 12:08:35
Job time: 10429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:36 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-18
Perfect score: 18
Sequence: 1 CTCGGTGAACGGCACCCC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87621	A87621 Sequence 18
2	c	18	100.0	1194	2	MTU92472	U92472 Mycobacteri
3	c	18	100.0	1208	5	A89752	A89752 Sequence 9
4	c	18	100.0	1209	5	A87612	A87612 Sequence 9
5	c	18	100.0	1228	5	A87608	A87608 Sequence 5
6	c	18	100.0	1228	5	A89748	A89748 Sequence 5
7	c	18	100.0	1229	5	A87610	A87610 Sequence 7
8	c	18	100.0	1229	5	A89750	A89750 Sequence 7
9	c	18	100.0	1235	5	A87606	A87606 Sequence 3
10	c	18	100.0	1235	5	A87609	A87609 Sequence 6
11	c	18	100.0	1235	5	A87611	A87611 Sequence 8
12	c	18	100.0	1235	5	A89746	A89746 Sequence 3
13	c	18	100.0	1235	5	A89749	A89749 Sequence 6
14	c	18	100.0	1235	5	A89751	A89751 Sequence 8
15	c	18	100.0	1236	5	A87613	A87613 Sequence 10
16	c	18	100.0	1236	5	A89753	A89753 Sequence 10
17	c	18	100.0	1237	5	A87607	A87607 Sequence 4
18	c	18	100.0	1237	5	A89747	A89747 Sequence 4
19	c	18	100.0	1245	5	A87605	A87605 Sequence 2
20	c	18	100.0	1245	5	A89745	A89745 Sequence 2
21	c	18	100.0	1260	5	A87604	A87604 Sequence 1
22	c	18	100.0	1260	5	A89744	A89744 Sequence 1
23	c	18	100.0	2584	1	MTALADH	X63069 M.tuberculo
24	c	18	100.0	56414	1	MTV002	AL008967 Mycobacte
25		16.4	91.1	10208	2	AE000795	AE000795 Methanoba
26	c	15.4	85.6	901	11	HSU16267	U16267 Human AMP d
27		15.4	85.6	1124	11	HSCOA8	AJ001544 Homo sapi
28	c	15.4	85.6	1129	11	HSU16268	U16268 Human AMP d
29	c	15.4	85.6	2422	10	HSU16269	U16269 Human AMP d
30	c	15.4	85.6	3359	9	S47833	S47833 AMPD2=AMP d
31	c	15.4	85.6	3386	9	HUMAMPD2	M91029 Human AMP d
32	c	15.4	85.6	11888	2	AF032114	AF032114 Methyloba
33	c	15.4	85.6	36225	1	MLCL581	Z96801 Mycobacteri
34		15.4	85.6	39991	1	MTCY274	Z74024 Mycobacteri
35		15.4	85.6	42210	1	SC1C2	AL031124 Streptomy
36	c	15.4	85.6	43454	40	AC002056	AC002056 Homo sapi
37	c	15.4	85.6	43885	40	AC000036	AC000036 Homo sapi
38		15.4	85.6	137972	43	AC011850	AC011850 Homo sapi
39		15.4	85.6	164549	55	AC023476	AC023476 Homo sapi
40		15.4	85.6	184235	44	AC012661	AC012661 Homo sapi
41		15.4	85.6	189370	2	AF010496	AF010496 Rhodobact
42		15	83.3	1549	4	U00676	U00676 Torpedo cal
43		15	83.3	1700	11	HSU40572	U40572 Human beta2
44		15	83.3	2261	1	MFTAPGENE	AJ000283 Mycobacte
45	c	15	83.3	2431	2	HIU17642	U17642 Haemophilus

ALIGNMENTS

```
RESULT   1
A87621
LOCUS       A87621         18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 18 from Patent WO9836089.
ACCESSION   A87621
VERSION     A87621.1  GI:6736261
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      8 c      5 g      2 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db      1 CTCGGTGAACGGCACCCC 18


RESULT   2
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1155 CTCGGTGAACGGCACCCC 1138


RESULT   3
A89752/c
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1155 CTCGGTGAACGGCACCCC 1138


RESULT   4
A87612/c
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1209;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1155 CTCGGTGAACGGCACCCC 1138

RESULT 5
A87608/c
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:6736248
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1155 CTCGGTGAACGGCACCCC 1138

RESULT 6
A89748/c
LOCUS A89748 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9832862.
ACCESSION A89748
VERSION A89748.1 GI:6738282
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1155 CTCGGTGAACGGCACCCC 1138

RESULT 7
A87610/c
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1154 CTCGGTGAACGGCACCCC 1137

RESULT 8
A89750/c
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9832862.
ACCESSION A89750
VERSION A89750.1 GI:6738284
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1154 CTCGGTGAACGGCACCCC 1137

RESULT 9
A87606/c
LOCUS A87606 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent WO9836089.

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1155 CTCGGTGAACGGCACCCC 1138


RESULT  10
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1154 CTCGGTGAACGGCACCCC 1137


RESULT  11
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1154 CTCGGTGAACGGCACCCC 1137


RESULT  12
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1155 CTCGGTGAACGGCACCCC 1138


RESULT  13
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||||||||||||||
Db    1154 CTCGGTGAACGGCACCCC 1137


RESULT  14
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||||||||||||||
Db    1154 CTCGGTGAACGGCACCCC 1137


RESULT  15
A87613/c
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||||||||||||||
Db    1155 CTCGGTGAACGGCACCCC 1138
```

Search completed: June 22, 2000, 14:58:38
Job time: 17933 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:54 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-18
Perfect score: 18
Sequence: 1 CTCGGTGAACGGCACCCC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49519	Mycobacterium sp.
2		18	100.0	18	1	V49617	AlaDH derived olig
3	c	18	100.0	1245	1	V49626	Mycobacterium tube
4	c	18	100.0	1260	1	V49510	Mycobacterium sp.
5	c	18	100.0	1260	1	V49625	Mycobacterium tube
6	c	14.8	82.2	1710	1	Q24075	Sarcosine oxidase.
7	c	14.8	82.2	2116	1	N50499	Pleiotropic regula
8		14.8	82.2	6151	1	Q75977	Insulin receptor s
9	c	14.8	82.2	15664	1	T09312	Mycobacteriophage
10	c	14.8	82.2	15664	1	T51224	NheI-G/SpeI fragme
11	c	14.8	82.2	15664	1	T66130	Mycobacteriophage
12	c	14.8	82.2	15664	1	T70492	Mycobacteriophage
13	c	14	77.8	37	1	T78194	Micro gene random
14	c	13.8	76.7	195	1	T00416	Genetic suppressor
15		13.8	76.7	582	1	N90494	DNA sequence encod
16		13.8	76.7	582	1	N90496	DNA sequence encod
17		13.8	76.7	815	1	Q10454	Adenylate kinase g
18	c	13.8	76.7	1125	1	V19021	Plasmid pPF1 struc
19		13.8	76.7	1233	1	T58557	Streptomyces prist
20	c	13.8	76.7	1269	1	T33617	DxrA gene encoding
21		13.8	76.7	1311	1	X15501	DNA encoding a his
22		13.8	76.7	1561	1	T58556	Streptomyces prist
23		13.8	76.7	1588	1	N98906	Inducible phytate
24	c	13.8	76.7	1931	1	X24407	Maize myo-inositol
25	c	13.8	76.7	2081	1	V01452	Modified doxA gene
26		13.8	76.7	2146	1	Q29239	Isocitrate dehydro
27	c	13.8	76.7	3013	1	V01451	Plasmid pANT195 in
28	c	13.8	76.7	3196	1	V01453	Streptomyces dauno
29	c	13.8	76.7	3196	1	V01447	Daunomycin C-14 hy
30	c	13.8	76.7	3546	1	X24410	Maize myo-inositol
31	c	13.8	76.7	3546	1	X24411	Maize myo-inositol
32		13.8	76.7	15872	1	T68715	Streptomyces venez
33		13.8	76.7	29879	1	Q46806	eryA region of S.
34		13.8	76.7	43280	1	T80413	Tylactone synthase
35	c	13.4	74.4	1950	1	V23479	Pseudomonas XcpQ s
36	c	13.4	74.4	2872	1	Q21001	Murine receptor-ty
37	c	13.4	74.4	287	1	T36881	Murine receptor ty
38	c	13.4	74.4	2872	1	X29853	Mouse receptor-typ
39		13.4	74.4	6000	1	Q86478	Human PTP-OB. Prot
40		13.4	74.4	6000	1	T85389	Human protein tyro
41		13.4	74.4	6000	1	X06095	Human protein tyro
42	c	13.4	74.4	7247	1	X20642	Polynucleotide seq
43		13.4	74.4	9515	1	Q55145	Pseudomonas aerugi
44		13.4	74.4	17612	1	V23494	Pseudomonas Xpc, O
45		13.4	74.4	110000	1	V30459_1	Continuation (2 of

ALIGNMENTS

RESULT 1
V49519
ID V49519 standard; DNA; 18 BP.
AC V49519;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R1.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 3 A; 8 C; 5 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1 CTCGGTGAACGGCACCCC 18

RESULT 2
V49617
ID V49617 standard; DNA; 18 BP.
AC V49617;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R1.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 3 A; 8 C; 5 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1 CTCGGTGAACGGCACCCC 18

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1161 CTCGGTGAACGGCACCCC 1144

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCCC 18
||||||||||||||||||
Db 1176 CTCGGTGAACGGCACCCC 1159

RESULT 5
V49625/c
ID V49625 standard; DNA; 1250 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match            100.0%;  Score 18;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 0.97;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||||||||||||||
Db   1176 CTCGGTGAACGGCACCCC 1159


RESULT   6
Q24075/c
ID   Q24075 standard; DNA; 1710 BP.
AC   Q24075;
DT   03-NOV-1992  (first entry)
DE   Sarcosine oxidase.
KW   Sarcosine oxidase; cell culture; ss.
OS   Actinomycetes.
FH   Key             Location/Qualifiers
FT   cds             165. .1331
FT                   /*tag=  a
FT                   /label= Sarcosine_oxidase
PN   J04094688-A.
PD   26-MAR-1992.
PF   10-AUG-1990; 213128.
PR   10-AUG-1990; JP-213128.
PA   (TOXN ) TOYO JOZO KK.
DR   WPI; 92-156058/19.
DR   P-PSDB; R23780.
PT   Actinomycetes derived sarcosine oxidase gene - prepd. by
PT   culturing Streptomyces lividans transformed by recombinant vector
PT   comprising Actinomycetes DNA
PS   Disclosure; Page 10; 14pp; Japanese.
CC   The sequence given encodes the Actinomycetes sarcosine oxidase gene.
CC   This DNA can be used to transform Streptomyces lividans which can
CC   then be cultured so that the gene is expressed and sarcosine oxidase
CC   can be collected from the medium.  This is an efficient way of
CC   producing sarcosine oxidase.  The strain of Streptomyces lividans
CC   pref. used is Tk24-pSOXS102.
SQ   Sequence   1710 BP;    245 A;    654 C;    578 G;    233 T;


  Query Match            82.2%;  Score 14.8;  DB 1;  Length 1710;
  Best Local Similarity  88.9%;  Pred. No. 43;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          | |||||| |||||||||
Db    818 CACGGTGATCGGCACCCC 801


RESULT   7
N50499/c
ID   N50499 standard; DNA; 2116 BP.
AC   N50499;
DT   20-JAN-1992  (first entry)
DE   Pleiotropic regulator gene afs B.
KW   Prodigiosin; actinohordine; A-factor.
OS   Stremptinomyces coelicolor.
FH   Key             Location/Qualifiers
FT   cds             532. .1740
FT                   /*tag=  a
FT                   /note= "Claim 1"
PN   J60012985-A.
PD   23-JAN-1985.
PF   05-JUL-1983; 122052.
PR   05-JUL-1983; JP-122052.
PA   (BEPP/) BEPPU T.
DR   WPI; 85-058092/10.
PT   Pleiotropic regulator gene - for distributed-feedback-type laser
PT   for forming diffracting-grid on curved semiconductor layer.
PT   NoAbstract Dwg 2b-d/2
PS   Disclosure; Fig 4; 5pp; Japanese.
CC   The gene is responsible for the production of A-factor, a minor
CC   regulator concerned with biosynthesis of streptomycin,
CC   autoresistance to the antibiotic and sporulation. The sequence is
CC   also concerned with the regulation of expression of the
CC   antimicrobial substances prodigiosin and actinorhodine.
SQ   Sequence   2116 BP;    314 A;    747 C;    780 G;    275 T;


  Query Match            82.2%;  Score 14.8;  DB 1;  Length 2116;
  Best Local Similarity  88.9%;  Pred. No. 43;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          |||||||||| || |||
Db    435 CTCGGTGAACGTCAGCCC 418


RESULT   8
Q75977
ID   Q75977 standard; DNA; 6151 BP.
AC   Q75977;
DT   20-JUL-1995  (first entry)
DE   Insulin receptor substrate-1.
KW   Insulin receptor substrate-1; IRS-1;
KW   non-insul n-dependent diabetes; NIDDM; ss.
OS   Homo sapi ns.
FH   Key             Location/Qualifiers
FT   cds             581. .4312
FT                   /*tag=  a
PN   WO9429345-A.
PD   22-DEC-1994.
PF   10-JUN-1994; DK0227.
PR   10-JUN-1993; DK-000683.
PR   09-AUG-1993; DK-000915.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Bjorbaek C,  Frederiksen KA,  Pedersen O;
DR   WPI; 95-036405/05.
DR   P-PSDB; R67708.
PT   DNA encoding mutated insulin receptor substrate 1 - and related
PT   vectors, transformed cells and mutant proteins, useful for
PT   detecting predisposition to non-insulin-dependent diabetes
PS   Disclosure; Page 31-38; 58pp; English.
CC   Genomic DNA from human leukocyte nuclei was subjected to PCR
CC   amplification using primers (given in Q75978-Q76015) specific for the
CC   human IRS-1 gene (Q75977).  Amplified DNA was subjected to SSCP
CC   analysis to detect mutations in the IRS-1 gene, indicating a
CC   predisposition to NIDDM.
SQ   Sequence   6151 BP;    1355 A;    1759 C;    1664 G;    1373 T;


  Query Match            82.2%;  Score 14.8;  DB 1;  Length 6151;
  Best Local Similarity  88.9%;  Pred. No. 44;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||| |||||| |||
Db   1603 CTCGGTGGACGGCAGCCC 1620


RESULT   9
T09312/c
ID   T09312 standard; DNA; 15664 BP.
AC   T09312;
```

```
DT   05-JUL-1996  (first entry)
DE   Mycobacteriophage genomic DS6A  15.6 Kb DNA fragment.
KW   MB DS6A; open reading frame; TB; Mycobacterium tuberculosis;
KW   TB complex species specific reporter mycobacterium assay;
KW   DNA polymerase; drug susceptibility; infection; ds.
OS   Mycobacteriophage DS6A.
FH   Key             Location/Qualifiers
FT   cds             4855. .5376
FT                   /*tag= a
FT                   /note= " encodes L5 gp37 homologue"
FT   misc_feature    222. .425
FT                   /*tag= b
FT                   /note= "potential open reading frame"
FT   misc_feature    451. .747
FT                   /*tag= c
FT                   /note= "potential open reading frame"
FT   misc_feature    747. .1109
FT                   /*tag= d
FT                   /note= "potential open reading frame"
FT   misc_feature    1109. .2014
FT                   /*tag= e
FT                   /note= "potential open reading frame"
FT   misc_feature    2034. .2747
FT                   /*tag= f
FT                   /note= "potential open reading frame"
FT   misc_feature    2747. .3109
FT                   /*tag= g
FT                   /note= "potential open reading frame"
FT   misc_feature    3109. .3444
FT                   /*tag= h
FT                   /note= "potential open reading frame"
FT   misc_feature    3444. .3728
FT                   /*tag= i
FT                   /note= "potential open reading frame"
FT   misc_feature    3731. .4855
FT                   /*tag= j
FT                   /note= "potential open reading frame"
FT   misc_feature    5382. .5747
FT                   /*tag= k
FT                   /note= "potential open readi.g frame"
FT   misc_feature    5837. .6307
FT                   /*tag= l
FT                   /note= "potential open reading frame"
FT   misc_feature    6403. .7770
FT                   /*tag= m
FT                   /note= "potential open reading frame"
FT   misc_feature    7770. .8006
FT                   /*tag= n
FT                   /note= "potential open reading frame"
FT   misc_feature    8033. .8236
FT                   /*tag= o
FT                   /note= "potential open reading frame"
FT   misc_feature    8244. .9443
FT                   /*tag= p
FT                   /note= "potential open reading frame"
FT   misc_feature    9450. .10244
FT                   /*tag= q
FT                   /note= "potential open reading frame"
FT   misc_feature    10371. .10586
FT                   /*tag= r
FT                   /note= "potential open reading frame"
FT   misc_feature    11115. .11786
FT                   /*tag= s
FT                   /note= "potential open reading frame"
FT   misc_feature    11917. .12741
FT                   /*tag= t
FT                   /note= "potential open reading frame"
FT   misc_feature    12748. .14499
FT                   /*tag= u
FT                   /note= "potential open reading frame"
FT   misc_feature    14771. .15154
FT                   /*tag= v
FT                   /note= "potential open reading frame"
FT   misc_feature    15154. .15426
FT                   /*tag= w
FT                   /note= "potential open reading frame"
FT   misc_feature    15429. .15664
FT                   /*tag= x
FT                   /note= "potential open reading frame"
PN   US5476768-A.
PD   19-DEC-1995.
PF   10-MAR-1995; 402282.
PR   10-MAR-1995; US-402282.
PA   (BECT ) BECTON DICKINSON CO.
PI   Beyer WF,  Dickson JA,  Hamilton PT,  Little MC,  Pearson RE;
DR   WPI; 96-048979/05.
PT   Myco:bacteriophage (MB) DS6A genomic DNA sequences - used for the
PT   rapid detection of TB complex MB species, and for determining MB
PT   drug susceptibility
PS   Claim 1; Column 25-39; 23pp; English.
CC   T09310-T09312 are genomic NheI fragments of mycobacteriophage DS6A.
CC   T09312 isa 15.6 Kb NheI fragment encoding an L5 gp37 gene homologue.
CC   DS6A is specific for mycobacterium tuberculosis (MB TB), it infects
CC   all species of the TB complex without any infection of other
CC   mycobacterium spp. The DS6A fragments can therefore be used in TB
CC   complex species specific reporter MB assays. The assays can be used
CC   to determine MB drug susceptibility.
SQ   Sequence   15664 BP;    2828 A;    5434 C;    5428 G;    1974 T;


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842


RESULT  10
T51224/c
ID   T51224 standard; DNA; 15664 BP.
AC   T51224;
DT   18-MAR-1997  (first entry)
DE   NheI-G/SpeI fragment of DS6A genome.
KW   DS6A phage; gp36; gp200; 55 kD protein; Mycobacterium; M. tuberculosis;
KW   TB complex; mycobacteriophage; M. bovis; M. africanum; M. bovis BCG;
KW   M. microti; reporter phage; ds.
OS   Mycobacteriophage DS6A.
FH   Key             Location/Qualifiers
FT   cds             222. .4258
FT                   /*tag= a
FT                   /note= "ORF 1"
FT   cds             451. .747
FT                   /*tag= b
FT                   /note= "ORF 2"
FT   cds             747. .1109
FT                   /*tag= c
FT                   /note= "ORF 3"
FT   cds             1109. .2014
FT                   /*tag= d
FT                   /note= "ORF 4"
FT   cds             2034. .2747
FT                   /*tag= e
FT                   /note= "ORF 5"
FT   cds             2747. .3109
FT                   /*tag= f
FT                   /note= "ORF 6"
FT   cds             3109. .3444
FT                   /*tag= g
FT                   /note= "ORF 7"
FT   cds             3444. .3728
FT                   /*tag= h
FT                   /note= "ORF 8"
FT   cds             3731. .4855
FT                   /*tag= i
```

```
FT                   /note= "ORF 9"
FT   cds             4855. .5376
FT                   /*tag= j
FT                   /note= "ORF 10 (L5 gp37 homologue)"
FT   cds             5382. .5747
FT                   /*tag= k
FT                   /note= "ORF 11"
FT   cds             5837. .6307
FT                   /*tag= l
FT                   /note= "ORF 12"
FT   cds             6403. .7770
FT                   /*tag= m
FT                   /note= "ORF 13"
FT   cds             7770. .8006
FT                   /*tag= n
FT                   /note= "ORF 14"
FT   cds             8033. .8236
FT                   /*tag= o
FT                   /note= "ORF 15"
FT   cds             8244. .9443
FT                   /*tag= p
FT                   /note= "ORF 16"
FT   cds             9450. .10244
FT                   /*tag= q
FT                   /note= "ORF 17"
FT   cds             10371. .10586
FT                   /*tag= r
FT                   /note= "ORF 18"
FT   cds             11115. .11786
FT                   /*tag= s
FT                   /note= "ORF 19"
FT   cds             11917. .12741
FT                   /*tag= t
FT                   /note= "ORF 20"
FT   cds             12748. .14499
FT                   /*tag= u
FT                   /note= "ORF 21"
FT   cds             14771. .15154
FT                   /*tag= v
FT                   /note= "ORF 22"
FT   cds             15154. .15426
FT                   /*tag= w
FT                   /note= "ORF 23"
FT   cds             15429. .15664
FT                   /*tag= x
FT                   /note= "ORF 24"
PN   US5582969-A.
PD   10-DEC-1996.
PF   10-MAR-1995; 402282.
PR   10-MAR-1995; US-402282.
PR   27-JUL-1995; US-508004.
PA   (BECT ) BECTON DICKINSON CO.
PI   Beyer WF,  Dickson JA,  Hamilton PT,  Little MC,  Pearson RE;
DR   WPI; 97-042294/04.
PT   Detection of TB complex mycobacteria - using labelled DS6A
PT   myco:bacteriophage
PS   Disclosure; Column 25-38; 22pp; English.
CC   T51222-T51224 represent fragments of the mycobacteriophage DS6A genome.
CC   This sequence represents the NheI-D fragment of DS6A.  The DS6A
CC   mycobacteriophage containing these sequences can be used in the method of
CC   the invention.  DS6A contains a 60 kD double stranded genome, and
CC   strongly resembles other mycobacteriophages in its physical
CC   characteristics.  DS6A is only lytic on mycobacteria contained within the
CC   TB complex, and does not lyse any other bacterial strain.  The method of
CC   the invention is for specifically detecting the presence of a TB complex
CC   mycobacterium.  The TB complex is the collective name for Mycobacterium
CC   tuberculosis and other mycobacteria closely related to it (M. bovis, M.
CC   africanum, M. bovis BCG, and M. microti).  The method comprises infecting
CC   a mycobacterium of the TB complex with a DS6A reporter mycobacteriophage,
CC   of a labelled DS6A phage, without infecting non-TB complex species.  A
CC   signal associated with the infected TB complex mycobacteria is then
CC   detected without detecting other mycobacterial species.  As DS6A infects
CC   all species of the TB complex without infecting other mycobacteria
CC   species, this diagnostic method eliminates the false positives of prior
CC   mycobacteriophage assays.
SQ   Sequence   15664 BP;    2828 A;    5434 C;    5428 G;    1974 T;


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842


RESULT  11
T66130/c
ID   T66130 standard; DNA; 15664 BP.
AC   T66130;
DT   10-JUN-1997  (first entry)
DE   Mycobacteriophage DS6A NheI-G/SpeI fragment.
KW   Mycobacterium; antibody; detection; TB complex; reporter phage DS6A;
KW   tuberculosis; glycoprotein; gp200; gp36; identification;
KW   mycobacterium bovis; africanum; microti; ds.
OS   Mycobacteriophage DS6A.
FH   Key             Location/Qualifiers
FT   misc_feature    222. .425
FT                   /*tag= a
FT                   /note= "potential open reading frame"
FT   misc_feature    451. .747
FT                   /*tag= b
FT                   /note= "potential open reading frame"
FT   misc_feature    747. .1109
FT                   /*tag= c
FT                   /note= "potential open reading frame"
FT   misc_feature    1109. .2014
FT                   /*tag= d
FT                   /note= "potential open reading frame"
FT   misc_feature    2034. .2747
FT                   /*tag= e
FT                   /note= "potential open reading frame"
FT   misc_feature    2747. .3109
FT                   /*tag= f
FT                   /note= "potential open reading frame"
FT   misc_feature    3109. .3444
FT                   /*tag= g
FT                   /note= "potential open reading frame"
FT   misc_feature    3444. .3728
FT                   /*tag= h
FT                   /note= "potential open reading frame"
FT   misc_feature    3731. .4855
FT                   /*tag= i
FT                   /note= "potential open reading frame"
FT   misc_feature    4855. .5376
FT                   /*tag= j
FT                   /note= "potential open reading frame"
FT   misc_feature    5382. .5747
FT                   /*tag= k
FT                   /note= "potential open reading frame"
FT   misc_feature    5837. .6307
FT                   /*tag= l
FT                   /note= "potential open reading frame"
FT   misc_feature    6403. .7770
FT                   /*tag= m
FT                   /note= "potential open reading frame"
FT   misc_feature    7770. .8006
FT                   /*tag= n
FT                   /note= "potential open reading frame"
FT   misc_feature    8033. .8236
FT                   /*tag= o
FT                   /note= "potential open reading frame"
FT   misc_feature    8244. .9443
FT                   /*tag= p
FT                   /note= "potential open reading frame"
```

```
FT   misc_feature    9450. .10244
FT                   /*tag= q
FT                   /note= "potential open reading frame"
FT   misc_feature    10371. .10586
FT                   /*tag= r
FT                   /note= "potential open reading frame"
FT   misc_feature    11115. .11786
FT                   /*tag= s
FT                   /note= "potential open reading frame"
FT   misc_feature    11917. .12741
FT                   /*tag= t
FT                   /note= "potential open reading frame"
FT   misc_feature    12748. .14499
FT                   /*tag= u
FT                   /note= "potential open reading frame"
FT   misc_feature    14771. .15154
FT                   /*tag= v
FT                   /note= "potential open reading frame"
FT   misc_feature    15154. .15426
FT                   /*tag= w
FT                   /note= "potential open reading frame"
FT   misc_feature    15429. .15664
FT                   /*tag= x
FT                   /note= "potential open read ng frame"
PN   US5612182-A.
PD   18-MAR-1997.
PF   10-MAR-1995; 402066.
PR   10-MAR-1995; US-402066.
PA   (BECT ) BECTON DICKINSON CO.
PI   Beyer WF,  Dickson JA,  Hamilton PT,  Little MC,  Pearson RE;
DR   WPI; 97-192104/17.
PT   Detection of mycobacteriophage DS6A - using labelled antibody to
PT   glyco:protein gp36 or gp200
PS   Disclosure; Column 25-38; 22pp; English.
CC   T66128-T66130 are fragments of the mycobacteriophage DS6A genome.
CC   The sequences can be used for the detection and identification of TB
CC   (tuberculosis) complex mycobacteria, e.g. M. tuberculosis 201, 303,
CC   -rif-r, 301, -ethamb-r, 509, 554, 1571, 1476, 790, M. bovis, M. bovis
CC   BCG, M. africanum and M. microti. Three structural proteins have
CC   been identified in DS6A: gp36, gp200 and a 55 kDA protein whose
CC   function is unknown. The N-terminal sequences of these proteins
CC   (see W11152-W11154) are used for the production of anti-gp36 and
CC   anti-gp200 antibodies which are used for the detection of DS6A. A
CC   reporter gene containing DS6A can be used for the specific detection
CC   of TB complex forming mycobacteria.
SQ   Sequence   15664 BP;    2828 A;    5434 C;    5428 G;    1974 T;

  Query Match           82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity  88.9%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCCC 18
          ||||||| ||||||| ||
Db   7859 CTCGGTGTACGGCACGCC 7842


RESULT  12
T70492/c
ID   T70492 standard; DNA; 15664 BP.
AC   T70492;
DT   05-JAN-1998  (first entry)
DE   Mycobacteriophage DS6A NheI-G/SpeI fragment.
KW   TB complex; tuberculosis; gp37; mycobacteriophage L5; DNA polymerase;
KW   beta subunit; Streptomyces coelicolor; reporter phage; detection;
KW   identification; NheI fragment D; N; G; SpeI; open reading frame;
KW   amplification; DNA synthesis; gp36; antibody; immunisation; antibody;
KW   immunoassay; ds.
OS   Mycobacteriphage DS6A.
FH   Key             Location/Qualifiers
FT   cds             222. .428
FT                   /*tag= a
FT                   /number= ORF_1
FT                   /note= "potential open reading frame"
FT   cds             451. .750
FT                   /*tag= b
FT                   /number= ORF_2
FT                   /note= "potential open reading frame"
FT   cds             747. .1112
FT                   /*tag= c
FT                   /number= ORF_3
FT                   /note= "potential open reading frame"
FT   cds             1109. .2017
FT                   /*tag= d
FT                   /number= ORF_4
FT                   /note= "potential open reading frame"
FT   cds             2034. .2750
FT                   /*tag= e
FT                   /number= ORF_5
FT                   /note= "potential open reading frame"
FT   cds             2747. .3112
FT                   /*tag= f
FT                   /number= ORF_6
FT                   /note= "potential open reading frame"
FT   cds             3109. .3447
FT                   /*tag= g
FT                   /number= ORF_7
FT                   /note= "potential open reading frame"
FT   cds             3444. .3731
FT                   /*tag= h
FT                   /number= ORF_8
FT                   /note= "potential open reading frame"
FT   cds             3731.4858
FT                   /*tag= i
FT                   /number= ORF_9
FT                   /note= "encodes DNA polymerase; claim 2"
FT   cds             4855. .5379
FT                   /*tag= j
FT                   /number= ORF_10
FT                   /product= L5_gp37_homologue
FT                   /note= "potential coding sequence"
FT   cds             5382. .5750
FT                   /*tag= k
FT                   /number= ORF_11
FT                   /note= "potential open reading frame"
FT   cds             5837. .6310
FT                   /*tag= l
FT                   /number= ORF_12
FT                   /note= "potential open reading frame"
FT   cds             6403. .7773
FT                   /*tag= m
FT                   /number= ORF_13
FT                   /note= "potential open reading frame"
FT   cds             7770. .8009
FT                   /*tag= n
FT                   /number= ORF_14
FT                   /note= "potential open reading frame"
FT   cds             8033. .8239
FT                   /*tag= o
FT                   /number= ORF_15
FT                   /note= "potential open reading frame"
FT   cds             8244. .9447
FT                   /*tag= p
FT                   /number= ORF_16
FT                   /note= "potential open reading frame"
FT   cds             9450. .10247
FT                   /*tag= q
FT                   /number= ORF_17
FT                   /note= "potential open reading frame"
FT   cds             10371. .10589
FT                   /*tag= r
FT                   /number= ORF_18
FT                   /note= "potential open reading frame"
FT   cds             11115. .11789
FT                   /*tag= s
FT                   /number= ORF_19
```

```
FT                   /note= "potential open reading frame"
FT   cds             11917. .12744
FT                   /*tag=  t
FT                   /number= ORF_20
FT                   /note= "potential open reading frame"
FT   cds             12748. .14502
FT                   /*tag=  u
FT                   /number= ORF_21
FT                   /note= "potential open reading frame"
FT   cds             14771. .15157
FT                   /*tag=  v
FT                   /number= ORF_22
FT                   /note= "potential open reading frame"
FT   cds             15154. .15429
FT                   /*tag=  w
FT                   /number= ORF_23
FT                   /note= "potential open reading frame"
FT   cds             15429. .15667
FT                   /*tag=  x
FT                   /number= ORF_24
FT                   /note= "potential open reading frame"
PN   US5633159-A.
PD   27-MAY-1997.
PF   10-MAR-1995; 402068.
PR   10-MAR-1995; US-402068.
PA   (BECT ) BECTON DICKINSON CO.
PI   Beyer WF,  Dickson JA,  Hamilton PT,  Little MC,  Pearson RE;
DR   WPI; 97-297364/27.
PT   DNA encoding Mycobacteriophage DS6A DNA polymerase III beta sub:unit
PT    - useful for nucleic acid amplification and DNA synthesis
PS   Disclosure; Column 23-38; 22pp; English.
CC   Mycobacteriophage DS6A has been characterised and found to specifically
CC   infect all species of the TB (tuberculosis) complex, without any
CC   detectable infection of mycobacteria species other than those of the TB
CC   complex.  DNA sequence analysis has revealed several potential open
CC   reading frames (ORF), including one encoding a protein analogous to gp37
CC   of mycobacteriophage L5 and a second encoding a protein with significant
CC   homology to the Streptomyces coelicolor DNA polymerase beta subunit.
CC   Based on the DNA sequence analysis, cloning sites can be identified for
CC   insertion of reporter genes, making DS6A useful as a reporter phage for
CC   specific detection and identification of species of the TB complex.
CC   Specifically the recombinant DNA polymerase III, beta subunit, of
CC   mycobacteriophage DS6A is claimed.  This is produced by expression of
CC   ORF 1 of the NheI D fragment (nucleotides 390-1538 of T70491). The class
CC   III DNA polymerase can be used in vitro in nucleic acid amplification
CC   and DNA synthesis. The protein encoded by ORF10 of the present sequence
CC   (which contains a 12 kb SpeI fragment plus the overlapping sequence of
CC   NheI fragment G) has homology to the L5 phage gp37 protein.
SQ   Sequence   15664 BP;    2828 A;    5434 C;    5428 G;    1974 T;
```

```
  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842
```

```
RESULT  13
T78194/c
ID   T78194 standard; DNA; 37 BP.
AC   T78194;
DT   09-OCT-1997  (first entry)
DE   Micro gene random polymer PCR primer KY-604.
KW   Polymerisation; Escherichia coli; isoleucyl tRNA synthase; ileS;
KW   polymerase chain reaction; ss.
OS   Synthetic.
PN   J09154585-A.
PD   17-JUN-1997.
PF   06-DEC-1995; 318396.
PR   06-DEC-1995; JP-318396.
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
DR   WPI; 97-367065/34.
PT   Preparation of a random polymer of a micro gene - using different
PT   specific DNA sequences attached to either end of the gene
PS   Example 1; Page 6; 8pp; Japanese.
CC   A method has been produced for the preparation of a random polymer of
CC   a microgene. The method involves adding a specific DNA sequence (A) to
CC   one end of a micro gene fragment and a different specific DNA sequence
CC   (B) to the other end, to generate DNA sequences (a) and (b), which
CC   contain partial sequences complementary to (A) and (B); and ligating
CC   single-stranded versions of the DNA sequences (a) and (b). The
CC   present sequence represents the PCR primer KY-604 which is used
CC   during the preparation of a random polymer of a micro gene. The
CC   method allows efficient polymerisation of the micro gene.
SQ   Sequence   37 BP;    6 A;    14 C;    14 G;    3 T;
```

```
  Query Match             77.8%;  Score 14;  DB 1;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 GTGAACGGCACCCC 18
           ||||||||||||||
Db      31 GTGAACGGCACCCC 18
```

```
RESULT  14
T00416/c
ID   T00416 standard; cDNA; 195 BP.
AC   T00416;
DT   26-MAR-1996  (first entry)
DE   Genetic suppressor element HL7.1.
KW   Genetic suppressor element; GSE; platinum-based drug; cisplatin;
KW   chemotherapy; HL7.1; testosterone-repressed prostatic message-2; TRPM-2;
KW   ss.
OS   Synthetic.
PN   WO9522612-A2.
PD   24-AUG-1995.
PF   22-FEB-1995; U02303.
PR   22-FEB-1994; US-199900.
PA   (UNII ) UNIV ILLINOIS FOUND.
PI   Gudkov A,  Kirschling DJ,  Roninson IB;
DR   WPI; 95-302718/39.
PT   Genetic suppressor elements which confer resistance to
PT   platinum-based drugs, eg. cisplatin, on cancer cells - useful for
PT   enhancement of chemotherapy, and for diagnosis of resistance to
PT   these drugs.
PS   Claim 14; Fig 17; 75pp; English.
CC   The sequences represented by T00405-T00418 are genetic suppressor
CC   elements (GSEs).  This sequence represents GSE HL6.10.  This sequence
CC   shows homology to the cDNA encoding testosterone-repressed prostatic
CC   message-2 (TRPM-2).  These sequences were obtained from a cDNA library
CC   derived from the total cDNA of a cisplatin sensitive cell.  Genetic
CC   suppressor elements confer resistance to platinum-based drugs (PDs), such
CC   as cisplatin.  These functional GSEs can then be used to create probes
CC   for the parent gene.  The probes can then be used in a method of
CC   measuring the level of GSE gene expression.  The GSEs can be used in
CC   methods of diagnosis of resistance to PDs by measuring the level of
CC   expression of GSE genes.  The GSEs are also used in methods to overcome
CC   resistance to PDs in cancer cells.  The GSEs (or fragments of them) can
CC   be used to inhibit the function of genes associated with sensitivity to
CC   PDs.  For enhancement of chemotherapy, a GSE can be transferred (either
CC   alone or with another gene) on an expression vector into blood progenitor
CC   cells from a cancer patient.  The cells are returned to the patients
CC   circulation and allowed to repopulate the blood before aggressive
CC   chemotherapy is carried out (using higher cisplatin concentrations than
CC   normal), this will thereby avoid toxic side effects to the immune system
CC   as the blood cells will be GSE resistant.
SQ   Sequence   195 BP;    46 A;    45 C;    71 G;    33 T;
```

```
  Query Match             76.7%;  Score 13.8;  DB 1;  Length 195;
  Best Local Similarity   88.2%;  Pred. No. 1.3e+02;
```

```
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           |||||| | ||||||||
Db      39 CTCGGTCAGCGGCACCC 23


RESULT  15
N90494
ID   N90494 standard; DNA; 582 BP.
AC   N90494;
DT   19-JUN-1990  (first entry)
DE   DNA sequence encoding swine adenylate kinase (SAK)
KW   Swine adenylate kinase; SAK.
OS   Sus scrofa.
FH   Key             Location/Qualifiers
FT   cds             1. .582
FT                   /*tag=  a
PN   J)1051087-A.
PD   27-FEB-1989.
PF   24-AUG-1987; 208332.
PR   24-AUG-1987; JP-208332.
PA   (NIHA) Nippon Mining KK.
PI
DR   WPI; 89-103996/14.
DR   P-PSDB; P93318.
PT   Expression vector for swine adenylate kinase -
PT   with N-terminal ATG before coding sequence, to increase
PT   expression
PS   Page 2; 5pp; Japanese.
CC   It is used in an expression vector for swine adenylate kinase (SAK). The
CC   expression vector contains this sequence with an N-terminal ATG before
CC   the Met codon. This increases the expression of SAK. The vector is pref.
CC   pBR322 or derivs.
SQ   Sequence   582 BP;   141 A;   154 C;   194 G;   93 T;


  Query Match            76.7%;  Score 13.8;  DB 1;  Length 582;
  Best Local Similarity  88.2%;  Pred. No. 1.4e 02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           ||||| ||| |||||||
Db      54 CTCGGGGAAGGGCACCC 70


Search completed: June 22, 2000, 15:12:57
Job time: 18103 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:22:01 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-18
Perfect score: 18
Sequence: 1 CTCGGTGAACGGCACCCC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.8	82.2	24	2	US-08-557-139-13	Sequence 13, Appl
	2	14.8	82.2	6152	2	US-08-557-139-1	Sequence 1, Appli
c	3	14.8	82.2	15664	1	US-08-402-282-3	Sequence 3, Appli
c	4	14.8	82.2	15664	1	US-08-508-004-3	Sequence 3, Appli
c	5	14.8	82.2	15664	1	US-08-402-066-3	Sequence 3, Appli
c	6	14.8	82.2	15664	1	US-08-402-068-3	Sequence 3, Appli
c	7	13.8	76.7	181	4	US-08-485-657A-19	Sequence 19, Appl
c	8	13.8	76.7	181	6	PCT-US95-02303-18	Sequence 18, Appl
c	9	13.8	76.7	195	4	US-08-485-657A-14	Sequence 14, Appl
c	10	13.8	76.7	195	6	PCT-US95-02303-14	Sequence 14, Appl
c	11	13.8	76.7	1125	3	US-08-709-874A-25	Sequence 25, Appl
c	12	13.8	76.7	1269	2	US-08-396-218-1	Sequence 1, Appli
c	13	13.8	76.7	1269	2	US-08-760-116-1	Sequence 1, Appli
	14	13.8	76.7	1311	3	US-08-899-028A-1	Sequence 1, Appli
	15	13.8	76.7	1311	4	US-09-210-124-1	Sequence 1, Appli
c	16	13.8	76.7	2081	4	US-09-096-982-7	Sequence 7, Appli
c	17	13.8	76.7	2081	4	US-08-653-650A-7	Sequence 7, Appli
c	18	13.8	76.7	3013	4	US-09-096-982-6	Sequence 6, Appli
c	19	13.8	76.7	3013	4	US-08-653-650A-6	Sequence 6, Appli
c	20	13.8	76.7	3196	4	US-09-096-982-4	Sequence 4, Appli
c	21	13.8	76.7	3196	4	US-08-653-650A-4	Sequence 4, Appli
	22	13.8	76.7	20235	2	US-07-642-734C-3	Sequence 3, Appli
	23	13.8	76.7	43280	3	US-08-804-227C-1	Sequence 1, Appli
c	24	13.4	74.4	2872	1	US-08-015-985-4	Sequence 4, Appli
	25	13.4	74.4	6000	1	US-08-348-006B-6	Sequence 6, Appli
	26	13.4	74.4	6000	3	US-08-800-825A-6	Sequence 6, Appli
	27	13.4	74.4	6000	6	PCT-US94-10166-6	Sequence 6, Appli
	28	13.4	74.4	9515	2	US-08-920-812-13	Sequence 13, Appl
	29	13.4	74.4	9515	2	US-08-920-827-13	Sequence 13, Appl
	30	13.4	74.4	9515	2	US-08-921-177-13	Sequence 13, Appl
	31	13.4	74.4	9515	2	US-08-362-577C-13	Sequence 13, Appl
	32	13.4	74.4	9515	3	US-08-920-828-13	Sequence 13, Appl
c	33	13.2	73.3	152	1	US-08-136-277-4	Sequence 4, Appli
c	34	13.2	73.3	152	3	US-08-479-403-4	Sequence 4, Appli
c	35	13.2	73.3	152	5	US-08-835-734-4	Sequence 4, Appli
c	36	13.2	73.3	354	4	US-08-474-379C-82	Sequence 82, Appl
	37	13.2	73.3	423	1	US-08-470-179-129	Sequence 129, App
	38	13.2	73.3	423	1	US-08-470-179-180	Sequence 180, App
c	39	13.2	73.3	760	4	US-08-474-379C-78	Sequence 78, Appl
	40	13.2	73.3	1345	2	US-08-525-654A-134	Sequence 134, App
c	41	13.2	73.3	1441	1	US-08-136-277-18	Sequence 18, Appl
c	42	13.2	73.3	1441	3	US-08-479-403-18	Sequence 18, Appl
c	43	13.2	73.3	1441	5	US-08-835-734-18	Sequence 18, Appl
	44	13.2	73.3	1528	4	US-08-945-296-3	Sequence 3, Appli
c	45	13.2	73.3	1672	1	US-07-688-352C-34	Sequence 34, Appl

ALIGNMENTS

```
RESULT   1
US-08-557-139-13/c
; Sequence 13, Application US/08557139
; Patent No. 5827730
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Oluf
;    APPLICANT:  Bjorbak, Christian
;    APPLICANT:  Frederiksen, Kathrine A.
;    TITLE OF INVENTION:  MUTANT DNA ENCODING INSULIN RECEPTOR
;    TITLE OF INVENTION:  SUBSTRATE 1
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5827730o No. 5827730disk of No. 5827730th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,139
;      FILING DATE:  12-FEB-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4041.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-557-139-13


  Query Match             82.2%;  Score 14.8;  DB 2;  Length 24;
  Best Local Similarity   88.9%;  Pred. No. 17;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| |||||| |||
```

```
Db      19 CTCGGTGGACGGCAGCCC 2


RESULT   2
US-08-557-139-1
; Sequence 1, Application US/08557139
; Patent No. 5827730
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Oluf
;    APPLICANT:  Bjorbak, Christian
;    APPLICANT:  Frederiksen, Kathrine A.
;    TITLE OF INVENTION:  MUTANT DNA ENCODING INSULIN RECEPTOR
;    TITLE OF INVENTION:  SUBSTRATE 1
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5827730o No. 5827730disk of No. 5827730th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,139
;      FILING DATE:  12-FEB-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4041.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6152 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  581..4309
US-08-557-139-1


  Query Match             82.2%;  Score 14.8;  DB 2;  Length 6152;
  Best Local Similarity   88.9%;  Pred. No. 23;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| |||||| |||
Db    1603 CTCGGTGGACGGCAGCCC 1620


RESULT   3
US-08-402-282-3/c
; Sequence 3, Application US/08402282
; Patent No. 5476768
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,282
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15664 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  222..425
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  451..747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  747..1109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1109..2014
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2034..2747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2747..3109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3109..3444
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3444..3728
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3731..4855
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
```

```
;      NAME/KEY:  misc_feature
;      LOCATION:  4855..5376
;      OTHER INFORMATION:  /function= "potential coding
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /product= "L5 gp37 homolog"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5382..5747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5837..6307
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6403..7770
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7770..8006
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8033..8236
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8244..9443
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9450..10244
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10371..10586
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11115..11786
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11917..12741
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12748..14499
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14771..15154
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15154..15426
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15429..15664
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-402-282-3


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 24;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           |||||||  |||||||  ||
Db    7859 CTCGGTGTACGGCACGCC 7842


RESULT   4
US-08-508-004-3/c
; Sequence 3, Application US/08508004
; Patent No. 5582969
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,004
;      FILING DATE:  27-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/402,282
;      FILING DATE:  10-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15664 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  222..425
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  451..747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  747..1109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
```

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1109..2014
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2034..2747
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2747..3109
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3109..3444
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3444..3728
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3731..4855
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4855..5376
;       OTHER INFORMATION:  /function= "potential coding
;       OTHER INFORMATION:  sequence"
;       OTHER INFORMATION:  /product= "L5 gp37 homolog"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5382..5747
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5837..6307
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6403..7770
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7770..8006
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8033..8236
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8244..9443
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9450..10244
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10371..10586
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11115..11786
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  11917..12741
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12748..14490
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14771..15154
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15154..15426
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15429..15664
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
US-08-508-004-3


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 24;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842


RESULT   5
US-08-402-066-3/c
; Sequence 3, Application US/08402066
; Patent No. 5612182
;  GENERAL INFORMATION:
;     APPLICANT:  Pearson, Robert E.
;     APPLICANT:  Dickson, Julie A.
;     APPLICANT:  Hamilton, Paul T.
;     APPLICANT:  Little, Michael C.
;     APPLICANT:  Beyer Jr., Wayne F.
;     TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;     TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;       ADDRESSEE:  Company
;       STREET:  1 Becton Drive
;       CITY:  Franklin Lakes
;       STATE:  NJ
;       COUNTRY:  US
;       ZIP:  07417
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/402,066
```

```
;      FILING DATE:
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15664 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  222..425
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  451..747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  747..1109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1109..2014
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2034..2747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2747..3109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3109..3444
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3444..3728
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3731..4855
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4855..5376
;      OTHER INFORMATION:  /function= "potential coding
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /product= "L5 gp37 homolog"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5382..5747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5837..6307
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6403..7770
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7770..8006
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8033..8236
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8244..9443
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9450..10244
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10371..10586
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11115..11786
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11917..12741
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12748..14499
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14771..15154
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15154..15426
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15429..15664
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-402-066-3


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 24;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842
```

```
RESULT   6
US-08-402-068-3/c
; Sequence 3, Application US/08402068
; Patent No. 5633159
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15664 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  222..425
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  451..747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  747..1109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1109..2014
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2034..2747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2747..3109
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3109..3444
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3444..3728
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3731..4855
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4855..5376
;      OTHER INFORMATION:  /function= "potential coding
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /product= "L5 gp37 homolog"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5382..5747
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5837..6307
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6403..7770
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7770..8006
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8033..8236
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8244..9443
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9450..10244
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10371..10586
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11115..11786
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  11917..12741
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12748..14499
;      OTHER INFORMATION:  /function= "potential open reading
```

```
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14771..15154
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15154..15426
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15429..15664
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-402-068-3


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 15664;
  Best Local Similarity   88.9%;  Pred. No. 24;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCCC 18
           ||||||| ||||||| ||
Db    7859 CTCGGTGTACGGCACGCC 7842


RESULT   7
US-08-485-657A-19/c
; Sequence 19, Application US/08485657A
; Patent No. 5942389
;  GENERAL INFORMATION:
;    APPLICANT:  Kirschling, Deborah J
;    APPLICANT:  Gudkov, Andrei
;    APPLICANT:  Roninson, Igor B
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive, 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,657A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5942389nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  93,354-N
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  181 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-485-657A-19


  Query Match             76.7%;  Score 13.8;  DB 4;  Length 181;
  Best Local Similarity   88.2%;  Pred. No. 62;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           |||||| | ||||||||
Db      25 CTCGGTCAGCGGCACCC 9


RESULT   8
PCT-US95-02303-18/c
; Sequence 18, Application PC/TUS9502303
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02303
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  181 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US95-02303-18


  Query Match             76.7%;  Score 13.8;  DB 6;  Length 181;
  Best Local Similarity   88.2%;  Pred. No. 62;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           |||||| | ||||||||
Db      25 CTCGGTCAGCGGCACCC 9


RESULT   9
US-08-485-657A-14/c
; Sequence 14, Application US/08485657A
; Patent No. 5942389
;  GENERAL INFORMATION:
;    APPLICANT:  Kirschling, Deborah J
;    APPLICANT:  Gudkov, Andrei
;    APPLICANT:  Roninson, Igor B
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive, 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,657A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5942389nan, Kevin E
;      REGISTRATION NUMBER:  35,303
```

```
;      REFERENCE/DOCKET NUMBER:  93,354-N
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-485-657A-14


  Query Match             76.7%;  Score 13.8;  DB 4;  Length 195;
  Best Local Similarity   88.2%;  Pred. No. 63;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCC 17
          ||||||  | ||||||||
Db     39 CTCGGTCAGCGGCACCC 23


RESULT  10
PCT-US95-02303-14/c
; Sequence 14, Application PC/TUS9502303
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Genes and Genetic Elements Associated
;    TITLE OF INVENTION:  with Sensitivity to Cisplatin
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02303
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US95-02303-14


  Query Match             76.7%;  Score 13.8;  DB 6;  Length 195;
  Best Local Similarity   88.2%;  Pred. No. 63;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCC 17
          ||||||  | ||||||||
Db     39 CTCGGTCAGCGGCACCC 23


RESULT  11
US-08-709-874A-25/c
; Sequence 25, Application US/08709874A
; Patent No. 5854040
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Chiba, Shigeru
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Trans-4-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS, LLP
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/709,874A
;      FILING DATE:  09-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/301,653
;      FILING DATE:  07-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/482,554
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terry, David T.
;      REGISTRATION NUMBER:  20178
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-312-6600
;      TELEFAX:  703-312-6666
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1125 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-709-874A-25


  Query Match             76.7%;  Score 13.8;  DB 3;  Length 1125;
  Best Local Similarity   88.2%;  Pred. No. 69;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCC 17
          | ||||||||||| ||
Db   1108 CACGGTGAACGGCAACC 1092


RESULT  12
US-08-396-218-1/c
; Sequence 1, Application US/08396218
; Patent No. 5695966
;  GENERAL INFORMATION:
;    APPLICANT:  INVENTI, Augusto
;    APPLICANT:  BREME, Umberto
;    APPLICANT:  COLOMBO, Anna L
;    APPLICANT:  HUTCHINSON, Charles R
;    APPLICANT:  OTTEN, Sharee
;    APPLICANT:  SCOTTI, Claudio
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIKAIDO, MARMELSTEIN, MURRAY & ORAM
;      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
;      STREET:  Street Lobby
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,218
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-5002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces peucetius
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1269
US-08-396-218-1


  Query Match             76.7%;  Score 13.8;  DB 2;  Length 1269;
  Best Local Similarity   88.2%;  Pred. No. 70;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TCGGTGAACGGCACCCC 18
           ||||||| |||||| ||
Db     518 TCGGTGACCGGCACACC 502


RESULT  13
US-08-760-116-1/c
; Sequence 1, Application US/08760116
; Patent No. 5786190
;  GENERAL INFORMATION:
;    APPLICANT:  INVENTI, Augusto
;    APPLICANT:  BREME, Umberto
;    APPLICANT:  COLOMBO, Anna L
;    APPLICANT:  HUTCHINSON, Charles R
;    APPLICANT:  OTTEN, Sharee
;    APPLICANT:  SCOTTI, Claudio
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIKAIDO, MARMELSTEIN, MURRAY & ORAM LLP
;      STREET:  655 Fifteenth Street, N. W., Suite 330 - G
;      STREET:  Street Lobby
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/760,116
;      FILING DATE:  3-DEC-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,218
;      FILING DATE:  27-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-6007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces peucetius
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1269
US-08-760-116-1


  Query Match             76.7%;  Score 13.8;  DB 2;  Length 1269;
  Best Local Similarity   88.2%;  Pred. No. 70;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TCGGTGAACGGCACCCC 18
           ||||||| |||||| ||
Db     518 TCGGTGACCGGCACACC 502


RESULT  14
US-08-899-028A-1
; Sequence 1, Application US/08899028A
; Patent No. 5858720
;  GENERAL INFORMATION:
;    APPLICANT:  Reichard, Raymond W.
;    APPLICANT:  Brown, James R.
;    APPLICANT:  Lawlor, Elizabeth J.
;    TITLE OF INVENTION:  No. 5858720el hisS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,028A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  GM10048
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1311 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

US-08-899-028A-1

Query Match 76.7%; Score 13.8; DB 3; Length 1311;
Best Local Similarity 88.2%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 TCGGTGAACGGCACCCC 18
           ||||||  |||||||||
Db     440 TCGGTGTGCGGCACCCC 456
```

RESULT 15
US-09-210-124-1
; Sequence 1, Application US/09210124
; Patent No. 5989884
; GENERAL INFORMATION:
; APPLICANT: Reichard, Raymond W.
; APPLICANT: Brown, James R.
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 5989884el hisS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,124
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/899,028
; FILING DATE: 07/23/1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10048
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-210-124-1

Query Match 76.7%; Score 13.8; DB 4; Length 1311;
Best Local Similarity 88.2%; Pred. No. 70;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 TCGGTGAACGGCACCCC 18
           ||||||  |||||||||
Db     440 TCGGTGTGCGGCACCCC 456
```

Search completed: June 22, 2000, 15:22:03
Job time: 18801 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:35 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-18
Perfect score: 18
Sequence: 1 CTCGGTGAACGGCACCCC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query %

	No.	Score	Match	Length	DB	ID	Description
c	1	15.4	85.6	402	20	D24758	D24758 RICR2515A R
c	2	15.4	85.6	426	79	AW301021	AW301021 xk11e05.x
c	3	15.4	85.6	438	51	AU070220	AU070220 AU070220
c	4	15.4	85.6	452	39	AA840367	AA840367 vw92c11.r
	5	15.4	85.6	492	106	AQ404964	AQ404964 HS_5048_B
c	6	15.4	85.6	512	64	AL119787	AL119787 DKFZp761F
c	7	15.4	85.6	737	40	AA984286	AA984286 am83e05.s
c	8	15.4	85.6	737	85	AQ688702	AQ688702 nbxb0078M
	9	15	83.3	79	37	AA731299	AA731299 nw68g05.s
	10	15	83.3	136	51	AI761024	AI761024 wi10c05.x
	11	15	83.3	146	40	AA974948	AA974948 on02b03.s
	12	15	83.3	178	34	AA535369	AA535369 nf94g03.s
	13	15	83.3	271	61	AI863542	AI863542 wh72g03.x
	14	15	83.3	278	45	AI334997	AI334997 tb21e01.x
	15	15	83.3	282	47	AI492904	AI492904 qz42e02.x
	16	15	83.3	318	40	AA954317	AA954317 on80f12.s
	17	15	83.3	325	39	AA832281	AA832281 oc91h07.s
	18	15	83.3	326	41	AI017917	AI017917 ou43e11.x
	19	15	83.3	330	39	AA831036	AA831036 oc58f02.s
	20	15	83.3	345	74	AW207063	AW207063 UI-H-BI1-
	21	15	83.3	347	37	AA731539	AA731539 nw63f09.s
	22	15	83.3	348	44	AI280452	AI280452 ql94g12.x
	23	15	83.3	354	38	AA742323	AA742323 ny68g06.s
	24	15	83.3	357	41	AI022872	AI022872 ow54d05.s
	25	15	83.3	362	39	AA830581	AA830581 oc55b09.s
	26	15	83.3	365	40	AA954307	AA954307 on80e12.s
	27	15	83.3	369	38	AA742382	AA742382 ny89e09.s
	28	15	83.3	369	38	AA748447	AA748447 ny01d06.s
	29	15	83.3	380	44	AI284272	AI284272 qi21g01.x
	30	15	83.3	408	31	AA279651	AA279651 zs87b03.s
	31	15	83.3	413	49	AI648366	AI648366 tx51h02.x
	32	15	83.3	415	43	AI217642	AI217642 qd70g01.x
	33	15	83.3	421	30	AA213648	AA213648 zr93c11.s
	34	15	83.3	425	60	AI809052	AI809052 wh74g09.x
	35	15	83.3	428	61	AI850261	AI850261 UI-M-BG1-
	36	15	83.3	429	49	AI627689	AI627689 ty81f11.x
	37	15	83.3	434	43	AI219837	AI219837 qg83d02.x
	38	15	83.3	439	74	AW207176	AW207176 UI-H-BI1-
	39	15	83.3	446	38	AA744130	AA744130 ny62b03.s
	40	15	83.3	450	31	AA278368	AA278368 zt08a01.s
	41	15	83.3	450	47	AI524154	AI524154 th09g11.x
	42	15	83.3	451	33	AA435658	AA435658 zt85h09.s
	43	15	83.3	452	64	AW028513	AW028513 wv27d11.x
	44	15	83.3	453	22	R47998	R47998 yj63g01.s1
	45	15	83.3	454	38	AA741303	AA741303 oc86b03.s

ALIGNMENTS

```
RESULT   1
D24758/c
LOCUS       D24758          402 bp    mRNA            EST       08-JUL-1999
DEFINITION  RICR2515A Rice root Oryza sativa cDNA clone R2515_1A, mRNA
            sequence.
ACCESSION   D24758
VERSION     D24758.1  GI:428606
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Minobe,Y. and Sasaki,T.
  TITLE     Rice cDNA from root
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R2515_1A"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT       89 a    111 c    100 g    102 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 20;  Length 402;
  Best Local Similarity   94.1%;  Pred. No. 3.3e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGGTGAACGGCACCC 17
          !!!!!!!!!! ||||||
Db     50 CTCGGTGAACTGCACCC 34


RESULT   2
AW301021/c
LOCUS       AW301021        426 bp    mRNA            EST       18-JAN-2000
DEFINITION  xk11e05.x1 NCI_CGAP_Co20 Homo sapiens cDNA clone IMAGE:2666432 3',
            mRNA sequence.
ACCESSION   AW301021
VERSION     AW301021.1  GI:6710698
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5422401.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2666432"
                     /clone_lib="NCI_CGAP_Co20"
                     /tissue_type="moderately differentiated adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Normalized to Cot 500.  Average insert size 1.11kb.
                     Normalized version of NCI_CGAP_Co18.  Library constructed
                     by Life Technologies."
BASE COUNT      112 a     65 c     97 g    115 t     37 others
```

```
ORIGIN

  Query Match             85.6%;  Score 15.4;  DB 79;  Length 426;
  Best Local Similarity   94.1%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           ||||||||| |||||||
Db     286 CTCGGTGAATGGCACCC 270


RESULT   3
AU070220/c
LOCUS       AU070220      438 bp    mRNA            EST       07-JUN-1999
DEFINITION  AU070220 Rice root Oryza sativa cDNA clone R3837_1A, mRNA sequence.
ACCESSION   AU070220
VERSION     AU070220.1  GI:5004982
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Minobe,Y. and Sasaki,T.
  TITLE     Rice cDNA from root
  JOURNAL   Unpublished (1995)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:31f 541.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R3837_1A"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT       88 a    130 c    129 g     84 t      7 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 51;  Length 438;
  Best Local Similarity   94.1%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           |||||||||| ||||||
Db     250 CTCGGTGAACTGCACCC 234


RESULT   4
AA840367/c
LOCUS       AA840367      452 bp    mRNA            EST       27-FEB-1998
DEFINITION  vw92c11.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:1262420 5' similar to TR:O00455 O00455 TTF-I INTERACTING
            PEPTIDE 20 ;, mRNA sequence.
ACCESSION   AA840367
VERSION     AA840367.1  GI:2916026
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151664.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:664972
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1262420"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       93 a    141 c    121 g     97 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 39;  Length 452;
  Best Local Similarity   94.1%;  Pred. No. 3.4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGGTGAACGGCACCC 17
           |||||||||| ||||||
Db     440 CTCGGTGAACCGCACCC 424


RESULT   5
AQ404964
LOCUS       AQ404964      492 bp    DNA             GSS       13-MAR-1999
DEFINITION  HS_5048_B2_C08_T7 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-624 Col-16 Row-F, genomic survey sequence.
ACCESSION   AQ404964
VERSION     AQ404964.1  GI:4410772
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     On Feb 19, 1999 this sequence version replaced gi:4142367.
```

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 624 row: F column: 16
Seq primer: T7
Class: BAC ends
High quality sequence stop: 492.

FEATURES Location/Qualifiers
source 1. .492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=624 Col=16 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT 123 a 136 c 89 g 144 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 106; Length 492;
Best Local Similarity 94.1%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCGGTGAACGGCACCC 17
||||||||| |||||||
Db 222 CTCGGTGAATGGCACCC 238

RESULT 6
AL119787/c
LOCUS AL119787 512 bp mRNA EST 27-SEP-1999
DEFINITION DKFZp761F2024_r1 761 (synonym: hamy2) Homo sapiens cDNA clone DKFZp761F2024 5', mRNA sequence.
ACCESSION AL119787
VERSION AL119787.1 GI:5925686
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 512)
AUTHORS Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jul 28, 1997 this sequence version replaced gi:2065403.
Contact: Bloecker H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by GBF within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES Location/Qualifiers
source 1. .512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761F2024"
/clone_lib="761 (synonym: hamy2)"
/tissue_type="amygdala"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 110 a 141 c 175 g 86 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 64; Length 512;
Best Local Similarity 94.1%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGGTGAACGGCACCCC 18
|| |||||||||||||
Db 358 TCTGTGAACGGCACCCC 342

RESULT 7
AA984286/c
LOCUS AA984286 737 bp mRNA EST 27-MAY-1998
DEFINITION am83e05.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:1629728 3' similar to gb:M91029 AMP DEAMINASE 2 (HUMAN);, mRNA sequence.
ACCESSION AA984286
VERSION AA984286.1 GI:3162811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404631.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop:.442.

FEATURES Location/Qualifiers
source 1. .737
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1529728"
/clone_lib="Stratagene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Yolken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of Individuals with Psychiatric Diseases (Unpublished) Stanley Neurovirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 154 a 211 c 219 g 149 t 4 others
ORIGIN

```
  Query Match             85.6%;  Score 15.4;  DB 40;  Length 737;
  Best Local Similarity   94.1%;  Pred. No. 4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TCGGTGAACGGCACCCC 18
          || |||||||||||||||
Db     92 TCTGTGAACGGCACCCC 76


RESULT   8
AQ688702/c
LOCUS       AQ688702      737 bp    DNA             GSS       01-JUL-1999
DEFINITION  nbxb0078M01r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0078M01r, genomic survey sequence.
ACCESSION   AQ688702
VERSION     AQ688702.1  GI:5329870
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta, Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     On Sep 10, 1998 this sequence version replaced gi:3553916.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence stop: 186.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0078M01r"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 26,864 clones with an average insert size
                     of 138.5 Kb providing 10.9 haploid genome equivalents.
                     The deep  coverage allows the isolation a particular
                     sequence with a probability  of 99.9 %. Two high density
                     filters, each containing 18,432 clones  (doubly spotted),
                     represent the whole library for colony screening."
BASE COUNT      184 a    204 c    186 g    161 t      2 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 85;  Length 737;
  Best Local Similarity   94.1%;  Pred. No. 4e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 TCGGTGAACGGCACCCC 18
          |||||||| |||||||||
Db    497 TCGGTGACCGGCACCCC 481


RESULT   9
AA731299
LOCUS       AA731299       79 bp    mRNA            EST       07-FEB-1998
DEFINITION  nw68g05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1251800 3'
            similar to TR:Q13425 Q13425 BETA2-SYNTROPHIN. ;, mRNA sequence.
ACCESSION   AA731299
VERSION     AA731299.1  GI:2752503
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 79)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394639.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered  erall poor quality
            Insert Length: 10    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .79
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1251800"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       17 a      30 c      24 g       8 t
ORIGIN


  Query Match             83.3%;  Score 15;  DB 37;  Length 79;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TCGGTGAACGGCACC 16
          |||||||||||||||
```

```
Db      20 TCGGTGAACGGCACC 34


RESULT  10
AI761024
LOCUS       AI761024      136 bp    mRNA            EST       24-JUN-1999
DEFINITION  wi10c05.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2389832 3'
            similar to TR:Q61235 Q61235 SYNTROPHIN-2. ;, mRNA sequence.
ACCESSION   AI761024
VERSION     AI761024.1  GI:5176691
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137968.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .136
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2389832"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       22 a     47 c     52 g     15 t
ORIGIN


  Query Match           83.3%;  Score 15;  DB 51;  Length 136;
  Best Local Similarity  100.0%;  Pred. No. 3.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TCGGTGAACGGCACC 16
          |||||||||||||||
Db     50 TCGGTGAACGGCACC 64


RESULT  11
AA974948
LOCUS       AA974948      146 bp    mRNA            EST       26-AUG-1998
DEFINITION  on02b03.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1555469 3'
            similar to TR:Q13425 Q13425 BETA2-SYNTROPHIN. ;, mRNA sequence.
ACCESSION   AA974948
VERSION     AA974948.1  GI:3150740
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:810962.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1231   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 55.
FEATURES             Location/Qualifiers
     source          1. .146
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1555469"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo. "
BASE COUNT       24 a     54 c     49 g     19 t
ORIGIN


  Query Match           83.3%;  Score 15;  DB 40;  Length 146;
  Best Local Similarity  100.0%;  Pred. No. 3.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TCGGTGAACGGCACC 16
          |||||||||||||||
Db     74 TCGGTGAACGGCACC 88


RESULT  12
AA535369
LOCUS       AA535369      178 bp    mRNA            EST       21-AUG-1997
DEFINITION  nf94g03.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:927604 3'
            similar to TR:G404933 G404933 SYNTROPHIN-2. ;, mRNA sequence.
ACCESSION   AA535369
VERSION     AA535369.1  GI:2279622
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 8, 1995 this sequence version replaced gi:801536.
```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 659 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .178
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:927604"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "

BASE COUNT 30 a 61 c 64 g 23 t
ORIGIN

Query Match 83.3%; Score 15; DB 34; Length 178;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGGTGAACGGCACC 16
||||||||||||||||
Db 92 TCGGTGAACGGCACC 106

RESULT 13
AI863542
LOCUS AI863542 271 bp mRNA EST 20-DEC-1999
DEFINITION wh72g03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2386324 3' similar to TR:Q61235 Q61235 SYNTROPHIN-2. ;, mRNA sequence.
ACCESSION AI863542
VERSION AI863542.1 GI:5527649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 271)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135715.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 546 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 262.

FEATURES Location/Qualifiers
source 1. .271
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2386324"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 35 a 92 c 111 g 33 t
ORIGIN

Query Match 83.3%; Score 15; DB 61; Length 271;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGGTGAACGGCACC 16
||||||||||||||||
Db 99 TCGGTGAACGGCACC 113

RESULT 14
AI334997
LOCUS AI334997 278 bp mRNA EST 29-DEC-1998
DEFINITION tb21e01.x1 NCI_CGAP_Kid12 Homo sapiens cDNA clone IMAGE:2055000 3' similar to TR:Q61235 Q61235 SYNTROPHIN, BASIC 2 ;, mRNA sequence.
ACCESSION AI334997
VERSION AI334997.1 GI:4071924
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 278)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151913.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2055000"
/clone_lib="NCI_CGAP_Kid12"

/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 67 a 71 c 79 g 61 t
ORIGIN

Query Match 83.3%; Score 15; DB 45; Length 278;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGGTGAACGGCACC 16
||||||||||||||||
Db 6 TCGGTGAACGGCACC 20

RESULT 15
AI492904
LOCUS AI492904 282 bp mRNA EST 17-MAR-1999
DEFINITION qz42e02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2029562 3' similar to TR:Q13425 Q13425 BETA2-SYNTROPHIN. ;, mRNA sequence.
ACCESSION AI492904
VERSION AI492904.1 GI:4393907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 282)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3136841.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1256 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 270.
FEATURES Location/Qualifiers
source 1. .282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2029562"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 57 a 85 c 87 g 53 t
ORIGIN

Query Match 83.3%; Score 15; DB 47; Length 282;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGGTGAACGGCACC 16
||||||||||||||||
Db 99 TCGGTGAACGGCACC 113

Search completed: June 22, 2000, 12:08:38
Job time: 10432 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:38 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-19
Perfect score: 18
Sequence: 1 GGCCAGCACGCTGGCGGG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87622	A87622 Sequence 19
2	c	18	100.0	1194	2	MTU92472	U92472 Mycobacteri
3	c	18	100.0	1208	5	A89752	A89752 Sequence 9
4	c	18	100.0	1209	5	A87612	A87612 Sequence 9
5	c	18	100.0	1228	5	A87608	A87608 Sequence 5
6	c	18	100.0	1228	5	A89748	A89748 Sequence 5
7	c	18	100.0	1229	5	A87610	A87610 Sequence 7
8	c	18	100.0	1229	5	A89750	A89750 Sequence 7
9	c	18	100.0	1235	5	A87606	A87606 Sequence 3
10	c	18	100.0	1235	5	A87609	A87609 Sequence 6
11	c	18	100.0	1235	5	A87611	A87611 Sequence 8
12	c	18	100.0	1235	5	A89746	A89746 Sequence 3
13	c	18	100.0	1235	5	A89749	A89749 Sequence 6
14	c	18	100.0	1235	5	A89751	A89751 Sequence 8
15	c	18	100.0	1236	5	A87613	A87613 Sequence 10
16	c	18	100.0	1236	5	A89753	A89753 Sequence 10
17	c	18	100.0	1237	5	A87607	A87607 Sequence 4
18	c	18	100.0	1237	5	A89747	A89747 Sequence 4
19	c	18	100.0	1245	5	A87605	A87605 Sequence 2
20	c	18	100.0	1245	5	A89745	A89745 Sequence 2
21	c	18	100.0	1260	5	A87604	A87604 Sequence 1
22	c	18	100.0	1260	5	A89744	A89744 Sequence 1
23	c	18	100.0	2584	1	MTALADH	X63069 M.tuberculo
24	c	18	100.0	56414	1	MTV002	AL008967 Mycobacte
25		17	94.4	2539	1	KPSCRR	X67751 K.pneumonia
26	c	16.4	91.1	439	7	FEN58SRR	L76775 Fennellia f
27		16.4	91.1	1413	12	D88497S09	D88505 Mouse DNA f
28		16.4	91.1	3566	12	MUSPALP	D43921 Mouse AZ1 m
29	c	16.4	91.1	4069	7	ZMWHPCS	X60204 Z.mays whp
30		16.4	91.1	10500	16	BHV1ULGNS	X94677 Bovine herp
31		16.4	91.1	37000	16	BHT1UL	Z78205 Bovine herp
32	c	16.4	91.1	42210	1	SC1C2	AL031124 Streptomy
33		16.4	91.1	43686	41	AC004396	AC004396 Homo sapi
34		16.4	91.1	73643	54	AC022983	AC022983 Homo sapi
35		16.4	91.1	135301	16	BHV1CGEN	AJ004801 Bovine he
36	c	16.4	91.1	149932	33	AL139423	AL139423 Homo sapi
37	c	16	88.9	2377	10	HSU27193	U27193 Human prote
38		16	88.9	3695	1	AB010722	AB010722 Rhodovulu
39		15.4	85.6	158	7	CA58SRRN	X06574 Cephalospor
40		15.4	85.6	175	7	YS8RGA1	M17093 Yeast (C.ac
41		15.4	85.6	333	8	AF122029	AF122029 Cordyceps
42	c	15.4	85.6	360	1	MNU17950	U17950 Mycobacteri
43	c	15.4	85.6	360	1	MNU17951	U17951 Mycobacteri
44	c	15.4	85.6	360	1	MSU17956	U17956 Mycobacteri
45		15.4	85.6	364	8	AF122027	AF122027 Cordyceps

ALIGNMENTS

```
RESULT   1
A87622
LOCUS       A87622         18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 19 from Patent WO9836089.
ACCESSION   A87622
VERSION     A87622.1  GI:6736262
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      6 c      9 g      1 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db      1 GGCCAGCACGCTGGCGGG 18


RESULT   2
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             PAT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1173 GGCCAGCACGCTGGCGGG 1156


RESULT   3
A89752/c
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1173 GGCCAGCACGCTGGCGGG 1156


RESULT   4
A87612/c
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1209;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1173 GGCCAGCACGCTGGCGGG 1156

RESULT 5
A87608/c
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:6736248
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1173 GGCCAGCACGCTGGCGGG 1156

RESULT 6
A89748/c
LOCUS A89748 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9832862.
ACCESSION A89748
VERSION A89748.1 GI:6738282
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1173 GGCCAGCACGCTGGCGGG 1156

RESULT 7
A87610/c
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1172 GGCCAGCACGCTGGCGGG 1155

RESULT 8
A89750/c
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9832862.
ACCESSION A89750
VERSION A89750.1 GI:6738284
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1172 GGCCAGCACGCTGGCGGG 1155

RESULT 9
A87606/c
LOCUS A87606 1235 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent WO9836089.

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1173 GGCCAGCACGCTGGCGGG 1156


RESULT  10
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1172 GGCCAGCACGCTGGCGGG 1155


RESULT  11
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1172 GGCCAGCACGCTGGCGGG 1155


RESULT  12
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1173 GGCCAGCACGCTGGCGGG 1156


RESULT  13
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1172 GGCCAGCACGCTGGCGGG 1155


RESULT  14
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1172 GGCCAGCACGCTGGCGGG 1155


RESULT  15
A87613/c
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1173 GGCCAGCACGCTGGCGGG 1156
```

```
Search completed: June 22, 2000, 14:58:40
Job time: 17935 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:57 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-19
Perfect score: 18
Sequence: 1 GGCCAGCACGCTGGCGGG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49520	Mycobacterium sp.
2		18	100.0	18	1	V49618	AlaDH derived olig
c 3		18	100.0	1245	1	V49626	Mycobacterium tube
c 4		18	100.0	1260	1	V49510	Mycobacterium sp.
c 5		18	100.0	1260	1	V49625	Mycobacterium tube
6		15.4	85.6	1254	1	T33127	Human neurotensin
7		15.4	85.6	2249	1	V84511	Human secreted pro
8		15	83.3	1302	1	T11308	Allergen Alt a 45
c 9		15	83.3	5199	1	X20508	Polynucleotide seq
10		14.8	82.2	39	1	Q37427	Primer -21NQ10BKNE
c 11		14.8	82.2	182	1	Q76564	Human genome fragm
12		14.8	82.2	553	1	V88883	EST clone HZ71. Ne
c 13		14.8	82.2	764	1	X19740	Human VRRP-1 (VR2),
14		14.8	82.2	831	1	Q99805	Thaumatin like gen
15		14.8	82.2	1400	1	Q91399	Pseudomonas protei
c 16		14.8	82.2	1809	1	T77791	DNA encoding 6-hyd
17		14.8	82.2	2679	1	Q28895	Fucose dehydrogena
c 18		14.8	82.2	2860	1	V59807	Human secreted pro
c 19		14.4	80.0	619	1	T70418	Mouse rhodopsin cD
c 20		14.4	80.0	811	1	V44446	Mycobacterium tube
c 21		14.4	80.0	811	1	V64555	M. tuberculosis im
22		14.4	80.0	825	1	T62585	CysE gene. Microbi
23		14.4	80.0	908	1	V68843	DNA molecule encod
24		14.4	80.0	1135	1	T94271	E. coli serine ace
25		14.4	80.0	1135	1	T94272	E. coli serine ace
26		14.4	80.0	1135	1	T94273	E. coli serine ace
27		14.4	80.0	1135	1	T94266	E. coli serine ace
28		14.4	80.0	1135	1	T94267	E. coli serine ace
29		14.4	80.0	1135	1	T94268	E. coli serine ace
30		14.4	80.0	1135	1	T94269	E. coli serine ace
31		14.4	80.0	1135	1	T94270	E. coli serine ace
32		14.4	80.0	1135	1	T94262	E. coli serine ace
33		14.4	80.0	1135	1	T94263	E. coli serine ace
34		14.4	80.0	1135	1	T94264	E. coli serine ace
35		14.4	80.0	1135	1	T94265	E. coli serine ace
36		14.4	80.0	1135	1	T94259	E. coli serine ace
37		14.4	80.0	1135	1	T94260	E. coli serine ace
38		14.4	80.0	1135	1	T94261	E. coli serine ace
39		14.4	80.0	1135	1	T66696	DNA encoding wild
40		14.4	80.0	1426	1	Q48396	Mycobacterium fort
c 41		14.4	80.0	1560	1	V63931	Mycobacterium tube
c 42		14.4	80.0	2471	1	Q55146	Pseudomonas aerugi
43		14.4	80.0	2847	1	Q43178	MTCE10 cysE expres
c 44		14.4	80.0	3807	1	V69215	MT-MMP-5 polypepti
45		14.4	80.0	5555	1	Q43180	MTCEK1 cassette fo

ALIGNMENTS

RESULT 1
V49520
ID V49520 standard; DNA; 18 BP.
AC V49520;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R2.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 2 A; 6 C; 9 G; 1 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1 GGCCAGCACGCTGGCGGG 18

RESULT 2
V49618
ID V49618 standard; DNA; 18 BP.
AC V49618;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R2.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 2 A; 6 C; 9 G; 1 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1 GGCCAGCACGCTGGCGGG 18

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent str ins, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 397 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1179 GGCCAGCACGCTGGCGGG 1162

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||||||||||
Db 1194 GGCCAGCACGCTGGCGGG 1177

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match           100.0%;  Score 18;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 4.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||||||||||||||
Db   1194 GGCCAGCACGCTGGCGGG 1177


RESULT   6
T33127
ID   T33127 standard; cDNA; 1254 BP.
AC   T33127;
DT   07-NOV-1996  (first entry)
DE   Human neurotensin receptor encoding sequence.
KW   neurotensin receptor; antagonist; agonist; screening; treatment; ulcer;
KW   Parkinson's disease; depression; dementia; retrograde oseophagitis; ds.
OS   Homo sapiens.
PN   J08143597-A.
PD   04-JUN-1996.
PF   24-NOV-1994; 289882.
PR   24-NOV-1994; JP-289882.
PA   (TAKE ) TAKEDA CHEM IND LTD.
DR   WPI; 96-318958/32.
DR   P-PSDB; R98562.
PT   DNA encoding human neurotensin receptor protein - useful for
PT   screening for (ant)agonists used to treat, e.g. Parkinson's disease,
PT   depression, dementia, retrograde oesophagitis, ulcers, etc.
PS   Claim 4; Page 22; 26pp; Japanese.
CC   The present sequence encodes a human neurotensin receptor protein
CC   (R98562).  The DNA and protein are useful for screening for
CC   (ant)agonists used to treat Parkinson's disease, depression, dementia,
CC   retrograde oseophagitis, ulcers, etc. .
SQ   Sequence   1254 BP;    215 A;    459 C;    359 G;    221 T;


  Query Match           85.6%;  Score 15.4;  DB 1;  Length 1254;
  Best Local Similarity  94.1%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          |||||||||| |||||
Db    646 GGCCAGCACGCCGGCGG 662


RESULT   7
V84511
ID   V84511 standard; DNA; 2249 BP.
AC   V84511;
DT   01-MAR-1999  (first entry)
DE   Human secreted protein gene 101 clone HSXBT86.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyrciditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9854963-A2.
PD   10-DEC-1998.
PF   04-JUN-1998; U11422.
PR   18-DEC-1997; US-070923.
PR   06-JUN-1997; US-048877.
PR   06-JUN-1997; US-048881.
PR   06-JUN-1997; US-048884.
PR   06-JUN-1997; US-048893.
PR   06-JUN-1997; US-048896.
PR   06-JUN-1997; US-048899.
PR   06-JUN-1997; US-048915.
PR   06-JUN-1997; US-048949.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048972.
PR   06-JUN-1997; US-049020.
PR   06-JUN-1997; US-049375.
PR   05-SEP-1997; US-057628.
PR   05-SEP-1997; US-057635.
PR   05-SEP-1997; US-057644.
PR   05-SEP-1997; US-057647.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057661.
PR   05-SEP-1997; US-057667.
PR   05-SEP-1997; US-057761.
PR   05-SEP-1997; US-057764.
PR   05-SEP-1997; US-057770.
PR   05-SEP-1997; US-057775.
PR   05-SEP-1997; US-057778.
PR   06-JUN-1997; US-048875.
PR   06-JUN-1997; US-048878.
PR   06-JUN-1997; US-048882.
PR   06-JUN-1997; US-048885.
PR   06-JUN-1997; US-048894.
PR   06-JUN-1997; US-048897.
PR   06-JUN-1997; US-048900.
PR   06-JUN-1997; US-048916.
PR   06-JUN-1997; US-048962.
PR   06-JUN-1997; US-048970.
PR   06-JUN-1997; US-048974.
PR   06-JUN-1997; US-049373.
PR   05-SEP-1997; US-057584.
PR   05-SEP-1997; US-057629.
PR   05-SEP-1997; US-057642.
PR   05-SEP-1997; US-057645.
PR   05-SEP-1997; US-057648.
PR   05-SEP-1997; US-057651.
PR   05-SEP-1997; US-057662.
PR   05-SEP-1997; US-057668.
PR   05-SEP-1997; US-057762.
PR   05-SEP-1997; US-057765.
PR   05-SEP-1997; US-057771.
PR   05-SEP-1997; US-057776.
PR   06-JUN-1997; US-048876.
PR   06-JUN-1997; US-048880.
PR   06-JUN-1997; US-048883.
PR   06-JUN-1997; US-048892.
PR   06-JUN-1997; US-048895.
PR   06-JUN-1997; US-048898.
PR   06-JUN-1997; US-048901.
PR   06-JUN-1997; US-048917.
PR   06-JUN-1997; US-048963.
PR   06-JUN-1997; US-048971.
PR   06-JUN-1997; US-049019.
PR   06-JUN-1997; US-049374.
PR   05-SEP-1997; US-057627.
PR   05-SEP-1997; US-057634.
PR   05-SEP-1997; US-057643.
PR   05-SEP-1997; US-057646.
PR   05-SEP-1997; US-057649.
PR   05-SEP-1997; US-057654.
PR   05-SEP-1997; US-057666.
PR   05-SEP-1997; US-057760.
PR   05-SEP-1997; US-057763.
PR   05-SEP-1997; US-057769.
PR   05-SEP-1997; US-057774.
```

PR 05-SEP-1997; US-057777.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
DR WPI; 99-059865/05.
DR P-PSDB; W88634, W88885, W88886.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 4; Page 362-364; 772pp; English.
CC The invention relates to nucleic acid sequences (V84411 to V84633)
CC encoding human secreted proteins (W88534 to W88756). The secreted protein
CC gene sequences are deposited with the ATCC under deposit numbers ATCC
CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
SQ Sequence 2249 BP; 524 A; 593 C; 581 G; 543 T;

Query Match 85.6%; Score 15.4; DB 1; Length 2249;
Best Local Similarity 94.1%; Pred. No. 66;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGG 17
|||||.|||||||||||
Db 284 GGCCATCACGCTGGCGG 300

RESULT 8
T11308
ID T11308 standard; cDNA to mRNA; 1302 BP.
AC T11308;
DT 19-AUG-1996 (first entry)
DE Allergen Alt a 45 protein encoding cDNA.
KW Alt a 45; Alt a 12; allergen; epitope; immunoglobulin E; IgE;
KW detection; immunotolerance; anergy; ds.
OS Alternaria alternata.
PN AT9402038-A.
PD 15-NOV-1995.
PF 02-NOV-1994; 002038.
PR 02-NOV-1994; AT-002038.
PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
PI Achatz G, Breitenbach M, Ebner C, Kraft D, Lechenauer E;
PI Oberkofler H, Prillinger H, Simon B, Unger A;
DR WPI; 96-040555/05.
DR P-PSDB; R88672.
PT Recombinant DNA encoding allergens of Alternaria alternata - useful
PT in diagnosis and treatment of A. alternata allergies
PS Claim 1; Page 8-10; 21pp; German.
CC The present sequence encodes an Alt a 45 allergen protein isolated from
CC Alternaria alternata. Peptide epitopes from Alt a 45 and Alt a 12
CC (R88692) are useful as diagnostic reagents, e.g. for in vitro detection
CC of allergy caused by Alt a 45 and 12 (by reaction with IgE in serum).
CC They can also detect cellular reaction to the specified allergens (from
CC their stimulatory or inhibitory effect on this reaction). The peptides
CC can also be used therapeutically to induce immunotolerance or anergy of
CC T lymphocytes.
SQ Sequence 1302 BP; 277 A; 433 C; 333 G; 259 T;

Query Match 83.3%; Score 15; DB 1; Length 1302;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGC 15
|||||||||||||||
Db 496 GGCCAGCACGCTGGC 510

RESULT 9
X20508/c
ID X20508 standard; DNA; 5199 BP.
AC X20508;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterization, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 227-230; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 5199 BP; 1167 A; 1492 C; 1492 G; 1256 T;

Query Match 83.3%; Score 15; DB 1; Length 5199;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GCCAGCACGCTGGCG 16
|||||||||||||||
Db 4500 GCCAGCACGCTGGCG 4486

RESULT 10
Q37427
ID Q37427 standard; DNA; 39 BP.
AC Q37427;
DT 19-JUN-1993 (first entry)
DE Primer -21NQ10BKNES.
KW Primer; heavy; light; chain; variable region; V; bcr-abl gene;
KW hybridoma; NQ2/12.4; NQ10/12.5; Vh; Vk; in-cell PCR; cloning;
KW polymorphic; Ig; TCR V; ss.
OS Synthetic.
PN WO9303151-A.
PD 18-FEB-1993.
PF 10-AUG-1992; G01483.
PR 10-AUG-1991; GB-017352.
PR 11-JUN-1992; GB-012419.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Embleton MJ, Gorochov G, Jones PT, Winter GP;

DR WPI; 93-076508/09.
PT Treatment of cell populations, partic. hybridomas - to link
PT together copies of 2 or more non-contiguous DNA sequences to
PT facilitate analysis
PS Disclosure; Page 27; 72pp; English.
CC The sequences given in Q37407-31 are primers which were based on
CC the sequence of the hybridoma V genes and the bcr-abl gene. These
CC primers were used for in-cell PCR to determine gene linkage analysis,
CC particularly for the cloning of gene combinations that are polymorphic
CC within a population of cells, such as the rearranged genes for Ig or
CC TCR V regions.
SQ Sequence 39 BP; 11 A; 10 C; 13 G; 5 T;

Query Match 82.2%; Score 14.8; DB 1; Length 39;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||| ||||||| |||
Db 12 GGCCAGTACGCTGGAGGG 29

RESULT 11
Q76564/c
ID Q76564 standard; DNA; 182 BP.
AC Q76564;
DT 23-SEP-1994 (first entry)
DE Human genome fragment.
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN WO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MEDI-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI; 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1; Page 98; 616pp; English.
CC Human nucleic acid fragments, isolated from brain adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q77613), (B) an allelic va. tion of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
SQ Sequence 182 BP; 59 A; 38 C; 45 G; 40 T;

Query Match 82.2%; Score 14.8; DB 1; Length 182;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||| | |||||||||
Db 127 GGCCAGAAAGCTGGCGGG 110

RESULT 12
V88883
ID V88883 standard; cDNA; 553 BP.
AC V88883;
DT 12-FEB-1999 (first entry)
DE EST clone HZ71.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN WO9845437-A2.
PD 15-OCT-1998.
PF 10-APR-1998; U06956.
PR 10-APR-1997; US-837312.
PA (GEMY) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 99-070078/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 544; 641pp; English.
CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequ ces isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 553 BP; 142 A; 164 C; 142 G; 104 T;

Query Match 82.2%; Score 14.8; DB 1; Length 553;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
|| ||| ||||||||||
Db 294 GGGCAGAACGCTGGCGGG 311

RESULT 13
X19740/c
ID X19740 standard; DNA; 764 BP.
AC X19740;
DT 16-JUN-1999 (first entry)
DE Human VRRP-1 (VR2) capsaicin receptor DNA sequence #2.
KW VR1; capsaicin receptor; VR2; VRRP-1; analgesic; diagnosis;
KW human disease; painful syndrome; ss.
OS Homo sapiens.
PN WO9909140-A1.
PD 25-FEB-1999.
PF 20-AUG-1998; U17466.
PR 22-JAN-1998; US-072151.
PR 20-AUG-1997; US-915461.
PA (REGC) UNIV CALIFORNIA.
PI Brake A, Caterina M, Julius DJ;
DR WPI; 99-181023/15.
PT New capsaicin receptor polypeptide - useful for screening or
PT characterising capsaicin receptor-binding compounds
PS Claim 8; Page 84-85; 99pp; English.
CC **The present sequence encodes an isolated capsaicin receptor polypeptide**
CC **(I). Capsaicin polypeptides are useful for identifying binding compounds**
CC which affect cellular responses. Preferably this is for identifying a
CC compound that binds (I) and affects a cellular response associated with
CC capsaicin biological activity (e.g. intracellular calcium flux). The
CC polypeptides and host cells are useful for detecting a vanilloid
CC **compound (an essential structural component of capsaicin) from natural**
CC products by detecting an alteration of intracellular response associated
CC with capsaicin receptor activity, preferably an alteration of
CC intracellular calcium levels, and are useful for screening for compounds
CC for use in analgesics. Capsaicin receptor polypeptides and antibodies
CC are useful for diagnosis and treatment of human diseases and painful
CC syndromes. The transgenic mammals can be used to screen for capsaicin
CC receptor antagonists and agonists. Prior art methods for screening or
CC characterising new capsaicin receptor-binding compounds relied on assays
CC using sensory neurons in culture or in intact animals. The new
CC polypeptides provide a more sensitive screen.

```
SQ   Sequence   764 BP;    163 A;    222 C;    202 G;    171 T;

  Query Match            82.2%;  Score 14.8;  DB 1;  Length 764;
  Best Local Similarity   88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          |||| ||| |||||||||
Db     98 GGCCTGCAGGCTGGCGGG 81


RESULT  14
Q99805
ID   Q99805 standard; cDNA; 831 BP.
AC   Q99805;
DT   20-JUN-1996  (first entry)
DE   Thaumatin like gene PR-5mz.
KW   SAR; tobacco; protein-synthesis independent gene; cyclohexamide;
KW   systemic acquired reistance response; anti-pathogen; plant protection;
KW   maize; PR-5; ss.
OS   Zea mays.
FH   Key             Location/Qualifiers
FT   cds             72. .590
FT                   /*tag= a
PN   WO9519443-A2.
PD   20-JUL-1995.
PF   03-JAN-1995; IB0002.
PR   13-JAN-1994; US-181271.
PA   (CIBA ) CIBA GEIGY AG.
PI   Alexander DC,  Ryals JA,  Uknes SJ,  Ward ER;
DR   WPI; 95-263872/34.
DR   P-PSDB; R81082.
PT   New DNA contg. plant systemic acquired resistance genes - and
PT   transgenic plants contg. them, impart disease and pest resistance,
PT   also Arabidopsis gene promoter to control DNA transcription
PS   Claim 21; Page 72-73; 85pp; English.
CC   This sequence represents the DNA sequence of a maize thaumatin like gene,
CC   PR-5mz.  This sequence was isolated by screening a BTH-induced cDNA
CC   library of maize, with a probe matching a known rice thaumatin clone.
CC   This sequence, and Q99800-Q99804 are all used in recombinant/chimaeric
CC   DNA molecules of the invention.  These sequences were isolated by
CC   differential screening of a cDNA library, followed by analysis by
CC   Northern hybridisation to RNA in the presence and absence of
CC   cyclohexamide.  The genes are used in the creation of transgenic plants.
CC   All of these sequences confer anti-pathogenic properties to transgenic
CC   plants.  Transgenic expression of 2 or more of the recombinant molecules
CC   of the invention that encode anti-pathogenic proteins provides a
CC   synergistic increase in plant protection, and may also offer protection
CC   against a wider range of pathogens.
SQ   Sequence   831 BP;    182 A;    259 C;    230 G;    160 T;

  Query Match            82.2%;  Score 14.8;  DB 1;  Length 831;
  Best Local Similarity   88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          |||| |||||||| ||||
Db    347 GGCCTGCACGCTGTCGGG 364


RESULT  15
Q91399
ID   Q91399 standard; DNA; 1400 BP.
AC   Q91399;
DT   07-FEB-1996  (first entry)
DE   Pseudomonas protein-activated lipase B gene.
KW   Protein activated lipase; lipB gene; ds.
OS   Pseudomonas aeruginosa.
FH   Key             Location/Qualifiers
FT   cds             38. .1057
FT                   /*tag= a
FT                   /product= lipase-B
PN   J07135971-A.
PD   30-MAY-1995.
PF   05-JUL-1993; 165561.
PR   05-JUL-1993; JP-165561.
PA   (TOYM ) TOYOBO KK.
DR   WPI; 95-227395/30.
DR   P-PSDB; R77317.
PT   A new protein-activated lipase - may be produced as a fusion
PT   protein with glutathione transferase, used in the production of
PT   lipase
PS   Claim 1; Page 5-6; 9pp; Japanese.
CC   Q91399 encodes a new protein-activated lipase. The protein is
CC   produced as a fusion protein with glutathionine transferase. The
CC   protein has an amino acid sequence of 339 residues and a molecular
CC   weight of 39 kD. The method used to produce the new protein yields
CC   active lipase of very high purity without using any complex
CC   purification process such as chromatography.
SQ   Sequence   1400 BP;    266 A;    504 C;    450 G;    180 T;

  Query Match            82.2%;  Score 14.8;  DB 1;  Length 1400;
  Best Local Similarity   88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          ||||||| ||| ||||||
Db    136 GGCCAGCCCGCAGGCGGG 153


Search completed: June 22, 2000, 15:12:59
Job time: 18105 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:22:03 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-19
Perfect score: 18
Sequence: 1 GGCCAGCACGCTGGCGGG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	14.8	82.2	39	3	US-08-190-199A-28	Sequence 28, Appl
	2	14.8	82.2	2682	1	US-07-855-793-3	Sequence 3, Appli
	3	14.4	80.0	1426	3	US-08-284-465-2	Sequence 2, Appli
	4	14.4	80.0	1426	3	US-08-284-465-7	Sequence 7, Appli
c	5	14.4	80.0	2471	2	US-08-920-812-14	Sequence 14, Appl
c	6	14.4	80.0	2471	2	US-08-920-827-14	Sequence 14, Appl
c	7	14.4	80.0	2471	2	US-08-921-177-14	Sequence 14, Appl
c	8	14.4	80.0	2471	2	US-08-362-577C-14	Sequence 14, Appl
c	9	14.4	80.0	2471	3	US-08-920-828-14	Sequence 14, Appl
c	10	14.4	80.0	3807	3	US-08-816-755-1	Sequence 1, Appli
	11	14.4	80.0	30001	1	US-08-125-468-1	Sequence 1, Appli
	12	14.4	80.0	30001	3	US-08-474-933-1	Sequence 1, Appli
c	13	14	77.8	2085	2	US-08-572-951-2	Sequence 2, Appli
c	14	14	77.8	5894	5	US-08-665-259-24	Sequence 2, Appl
c	15	13.8	76.7	268	1	US-08-105-168B-1	Sequence 1, Appli
c	16	13.8	76.7	268	1	US-08-105-168B-2	Sequence 2, Appli
c	17	13.8	76.7	268	1	US-08-105-168B-3	Sequence 3, Appli
c	18	13.8	76.7	268	1	US-08-105-168B-4	Sequence 4, Appli
c	19	13.8	76.7	268	3	US-08-698-948-1	Sequence 1, Appli
c	20	13.8	76.7	268	3	US-08-698-948-2	Sequence 2, Appli
c	21	13.8	76.7	268	3	US-08-698-948-3	Sequence 3, Appli
c	22	13.8	76.7	268	3	US-08-698-948-4	Sequence 4, Appli
c	23	13.8	76.7	314	1	US-08-105-168B-23	Sequence 23, Appl
c	24	13.8	76.7	314	3	US-08-698-948-23	Sequence 23, Appl
	25	13.8	76.7	330	6	PCT-US95-09589-1	Sequence 1, Appli
	26	13.8	76.7	330	6	PCT-US95-09589A-1	Sequence 1, Appli
c	27	13.8	76.7	342	3	US-08-473-020A-31	Sequence 31, Appl
c	28	13.8	76.7	343	3	US-08-473-020A-4	Sequence 4, Appli
c	29	13.8	76.7	343	3	US-08-473-020A-8	Sequence 8, Appli
c	30	13.8	76.7	350	1	US-08-105-168B-22	Sequence 22, Appl
c	31	13.8	76.7	350	3	US-08-698-948-22	Sequence 22, Appl
c	32	13.8	76.7	423	1	US-08-470-179-137	Sequence 137, App
c	33	13.8	76.7	442	4	US-08-949-584A-1	Sequence 1, Appli
	34	13.8	76.7	545	4	US-08-722-187-85	Sequence 85, Appl
c	35	13.8	76.7	545	4	US-08-722-187-85	Sequence 85, Appl
	36	13.8	76.7	545	6	PCT-US95-04712-85	Sequence 85, Appl
c	37	13.8	76.7	545	6	PCT-US95-04712-85	Sequence 85, Appl
	38	13.8	76.7	624	4	US-08-426-599B-1	Sequence 1, Appli
	39	13.8	76.7	852	2	US-08-728-259A-2	Sequence 2, Appli
	40	13.8	76.7	852	2	US-08-728-259A-3	Sequence 3, Appli
	41	13.8	76.7	852	2	US-08-728-259A-5	Sequence 5, Appli
	42	13.8	76.7	852	2	US-08-728-259A-7	Sequence 7, Appli
	43	13.8	76.7	852	4	US-08-473-486-2	Sequence 2, Appli
	44	13.8	76.7	852	4	US-08-473-486-3	Sequence 3, Appli
	45	13.8	76.7	852	4	US-08-473-486-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT   1
US-08-190-199A-28
; Sequence 28, Application US/08190199A
; Patent No. 5830663
;  GENERAL INFORMATION:
;    APPLICANT:  EMBLETON, Michael J.
;    APPLICANT:  GOROCHOV, Guy
;    APPLICANT:  JONES, Peter T.
;    APPLICANT:  WINTER, Gregory P.
;    TITLE OF INVENTION:  TREATMENT OF CELL POPULATIONS
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY MADISON & SUTRO, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/190,199A
;      FILING DATE:  13-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB92/01483
;      FILING DATE:  10-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9212419.7
;      FILING DATE:  11-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9117352.6
;      FILING DATE:  10-AUG-1991
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  39 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-190-199A-28


  Query Match             82.2%;  Score 14.8;  DB 3;  Length 39;
  Best Local Similarity   88.9%;  Pred. No. 86;
```

```
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCGGG 18
           |||||| ||||||| |||
Db      12 GGCCAGTACGCTGGAGGG 29


RESULT   2
US-07-855-793-3
; Sequence 3, Application US/07855793
; Patent No. 5217880
;  GENERAL INFORMATION:
;    APPLICANT:  Masanori MITTA et al.
;    TITLE OF INVENTION:  L-FUCOSE DEHYDROGENASE GENE,
;    TITLE OF INVENTION:  MICROORGANISM HAVING SAID GENE AND PRODUCTION OF L-FUCOSE
;    TITLE OF INVENTION:  DEHYDROGENASE BY THE USE OF SAID MICROORGANISM
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 Kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  DisplayWrite
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/855,793
;      FILING DATE:  19920323
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2682 Base Pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Arthrobacter Oxidans
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:  (A)  NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  /note= "844-1809 E CDS"
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-07-855-793-3


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 2682;
  Best Local Similarity   88.9%;  Pred. No. 77;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCGGG 18
           |||| |||||||||| ||
Db     167 GGCCTGCACGCTGGCTGG 184


RESULT   3
US-08-284-465-2
; Sequence 2, Application US/08284465
; Patent No. 5830457
;  GENERAL INFORMATION:
;    APPLICANT:  GICQUEL, BRIGITTE
;    APPLICANT:  TIMM, JULIANO
;    APPLICANT:  TRIAS, JOAQUIM
;    APPLICANT:  DUEZ, COLETTE
;    APPLICANT:  PERILLI, MARIA-GRAZIA
;    APPLICANT:  DUSART, JEAN
;    APPLICANT:  FRERE, JEAN-MARIE
;    TITLE OF INVENTION:  RECOMBINANT BETA-LACTAMASE, USABLE AS
;    TITLE OF INVENTION:  CARRIER MOLECULE FOR THE PREPARATION OF IMMUNOGENIC
;    TITLE OF INVENTION:  COMPOSITIONS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/284,465
;      FILING DATE:  14-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 9201713
;      FILING DATE:  14-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00151
;      FILING DATE:  12-FEB-1993
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1426 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  395..1276
US-08-284-465-2


  Query Match             80.0%;  Score 14.4;  DB 3;  Length 1426;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCAGCACGCTGGCGGG 18
           |||||||||  ||||||
Db     948 CCAGCACGCCGGCGGG 963


RESULT   4
US-08-284-465-7
; Sequence 7, Application US/08284465
; Patent No. 5830457
;  GENERAL INFORMATION:
;    APPLICANT:  GICQUEL, BRIGITTE
;    APPLICANT:  TIMM, JULIANO
;    APPLICANT:  TRIAS, JOAQUIM
;    APPLICANT:  DUEZ, COLETTE
;    APPLICANT:  PERILLI, MARIA-GRAZIA
;    APPLICANT:  DUSART, JEAN
;    APPLICANT:  FRERE, JEAN-MARIE
;    TITLE OF INVENTION:  RECOMBINANT BETA-LACTAMASE, USABLE AS
;    TITLE OF INVENTION:  CARRIER MOLECULE FOR THE PREPARATION OF IMMUNOGENIC
;    TITLE OF INVENTION:  COMPOSITIONS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/284,465
;      FILING DATE:  14-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 9201713
;      FILING DATE:  14-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00151
;      FILING DATE:  12-FEB-1993
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1426 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  395..1276
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  563..568
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  672..677
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  506..532
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  539..562
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  572..589
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  620..652
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  725..748
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  764..790
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  815..835
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  842..874
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  905..925
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  956..988
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1016..1036
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1052..1075
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1061..1066
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1091..1117
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1127..1156
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1172..1204
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1235..1276
US-08-284-465-7


  Query Match             80.0%;  Score 14.4;  DB 3;  Length 1426;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCAGCACGCTGGCGGG 18
           |||||||||  ||||||
Db     948 CCAGCACGCCGGCGGG 963


RESULT   5
US-08-920-812-14/c
; Sequence 14, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
```

```
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-812-14


  Query Match           80.0%;  Score 14.4;  DB 2;  Length 2471;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCAGCACGCTGGCGGG 18
           || ||||||||||||||
Db    1148 CCCGCACGCTGGCGGG 1133


RESULT   6
US-08-920-827-14/c
; Sequence 14, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-827-14


  Query Match           80.0%;  Score 14.4;  DB 2;  Length 2471;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCAGCACGCTGGCGGG 18
           || ||||||||||||||
Db    1148 CCCGCACGCTGGCGGG 1133


RESULT   7
US-08-921-177-14/c
; Sequence 14, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-921-177-14


  Query Match             80.0%;  Score 14.4;  DB 2;  Length 2471;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 CCAGCACGCTGGCGGG 18
          || |||||||||||||
Db   1148 CCCGCACGCTGGCGGG 1133


RESULT   8
US-08-362-577C-14/c
; Sequence 14, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-362-577C-14
```

```
  Query Match             80.0%;  Score 14.4;  DB 2;  Length 2471;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 CCAGCACGCTGGCGGG 18
          || |||||||||||||
Db   1148 CCCGCACGCTGGCGGG 1133


RESULT   9
US-08-920-828-14/c
; Sequence 14, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
;      STRAIN:  Clinical Isolate P2-7
US-08-920-828-14


  Query Match             80.0%;  Score 14.4;  DB 3;  Length 2471;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 CCAGCACGCTGGCGGG 18
          || |||||||||||||
Db   1148 CCCGCACGCTGGCGGG 1133


RESULT  10
US-08-816-755-1/c
; Sequence 1, Application US/08816755
```

```
; Patent No. 5837508
; GENERAL INFORMATION:
;   APPLICANT: Arleth, Anthony
;   APPLICANT: Arnold, Anne
;   APPLICANT: Shabon, Usman
;   TITLE OF INVENTION: No. 5837508el Membrane-Type Matrix
;   TITLE OF INVENTION: Metalloproteinase-5 Gene
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SmithKline Beecham Corporation
;     STREET: 709 Swedeland Road
;     CITY: King of Prussia
;     STATE: PA
;     COUNTRY:
;     ZIP: 19406
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/816,755
;     FILING DATE: 05-MAR-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Han, William T
;     REGISTRATION NUMBER: 34,344
;     REFERENCE/DOCKET NUMBER: GH50007
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 610-270-5219
;     TELEFAX: 610-270-4026
;     TELEX:
;  INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3807 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-816-755-1

  Query Match             80.0%;  Score 14.4;  DB 3;  Length 3807;
  Best Local Similarity   93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCG 16
           ||||||||||||||| |
Db    2853 GGCCAGCACGCTGGAG 2838


RESULT 11
US-08-125-468-1
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
;   APPLICANT: Ryan, Michael J.
;   APPLICANT: Lotvin, Jason A.
;   APPLICANT: Strathy, Nancy
;   APPLICANT: Fantini, Susan E.
;   TITLE OF INVENTION: Cloning of the biosynthetic pathway for
;   TITLE OF INVENTION: chlortetracycline and tetracyline Formation and cosmids
;   TITLE OF INVENTION: useful therein
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: American Cyanamid Company
;     STREET: One Cyanamid Plaza
;     CITY: Wayne
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07470
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/125,468
;     FILING DATE: 22-SEP-1993
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Tsevdos, Estelle J
;     REGISTRATION NUMBER: 31,145
;     REFERENCE/DOCKET NUMBER: 31,255-02
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (201)831-3241
;     TELEFAX: (201)831-3305
;  INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 30001 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

  Query Match             80.0%;  Score 14.4;  DB 1;  Length 30001;
  Best Local Similarity   93.8%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCCAGCACGCTGGCGG 17
           |||||| |||||||||
Db   24627 GCCAGCGCGCTGGCGG 24642


RESULT 12
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
;   APPLICANT: Ryan, Michael J.
;   APPLICANT: Lotvin, Jason A.
;   APPLICANT: Strathy, Nancy
;   APPLICANT: Fantini, Susan E.
;   TITLE OF INVENTION: Cloning of the biosynthetic pathway for
;   TITLE OF INVENTION: chlortetracycline and tetracyline Formation and cosmids
;   TITLE OF INVENTION: useful therein
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: American Cyanamid Company
;     STREET: One Cyanamid Plaza
;     CITY: Wayne
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07470
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/474,933
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/125,468
;     FILING DATE: 22-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Tsevdos, Estelle J
;     REGISTRATION NUMBER: 31,145
;     REFERENCE/DOCKET NUMBER: 31,255-02
;   TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-933-1


  Query Match             80.0%;  Score 14.4;  DB 3;  Length 30001;
  Best Local Similarity   93.8%;  Pred. No. 1.1e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCCAGCACGCTGGCGG 17
           |||||| |||||||||
Db   24627 GCCAGCGCGCTGGCGG 24642


RESULT  13
US-08-572-951-2/c
; Sequence 2, Application US/08572951
; Patent No. 5824790
;  GENERAL INFORMATION:
;    APPLICANT:  KEELING, PETER L.
;    APPLICANT:  KNIGHT, MARY E.
;    APPLICANT:  GUAN, HANPING
;    TITLE OF INVENTION:  MODIFICATION OF STARCH
;    TITLE OF INVENTION:  SYNTHESIS IN PLANTS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      ADDRESSEE:  Intellectual Property Group of
;      ADDRESSEE:  Pillsbury Madison & Sutro LLP
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/572,951
;      FILING DATE:  15-DEC-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/346,602
;      FILING DATE:  29-NOV-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/263,921
;      FILING DATE:  21-JUN-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul N. Kokulis
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  222957/1.02.15C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2085 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-572-951-2


  Query Match             77.8%;  Score 14;  DB 2;  Length 2085;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CCAGCACGCTGGCG 16
           ||||||||||||||
Db   1508 CCAGCACGCTGGCG 1495


RESULT  14
US-08-665-259-24/c
; Sequence 24, Application US/08665259
; Patent No. 6028173
;  GENERAL INFORMATION:
;    APPLICANT:  Landes, Gregory M.
;    APPLICANT:  Burn, Timothy C.
;    APPLICANT:  Connors, Timothy D.
;    APPLICANT:  Dackowski, William R.
;    APPLICANT:  Van Raay, Terence J.
;    APPLICANT:  Klinger, Katherine W.
;    TITLE OF INVENTION:  NOVEL HUMAN CHROMOSOME 16 GENES,
;    TITLE OF INVENTION:  COMPOSITIONS, METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  One Mountain Road
;      CITY:  Framingham
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,259
;      FILING DATE:  17-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dugan, Deborah A.
;      REGISTRATION NUMBER:  37,315
;      REFERENCE/DOCKET NUMBER:  IG5-9.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (508) 872-8400
;      TELEFAX:  (508) 872-5415
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5894 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..5053
US-08-665-259-24


  Query Match             77.8%;  Score 14;  DB 5;  Length 5894;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGG 14
           ||||||||||||||
Db    334 GGCCAGCACGCTGG 321


RESULT  15
US-08-105-168B-1/c
; Sequence 1, Application US/08105168B
```

```
; Patent No. 5589585
; GENERAL INFORMATION:
;   APPLICANT:  MABILAT et al.
;   TITLE OF INVENTION:  DNA FRAGMENTS OF MYCOBACTERIA, AMPLIFICATION
;   TITLE OF INVENTION:  PRIMERS, HYBRIDIZATION PROBES, REAGENTS AND METHOD FOR THE DET
;   TITLE OF INVENTION:  MYCOBACTERIA
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Oliff & Berridge
;     STREET:  700 South Washington Street, Suite 300
;     CITY:  Alexandria,
;     STATE:  Virginia
;     ZIP:  22314
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5" DS/HD
;     COMPUTER:  IBM compatible
;     OPERATING SYSTEM:  MS DOS 3.1
;     SOFTWARE:  Wordperfect
;   CURRENT APPLICATION DATA:
;     FILING DATE:  August 12, 1993
;     APPLICATION NUMBER:  US/08/105,168B
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  FR9210094
;     FILING DATE:  August 8, 1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  William P. Berridge
;     REGISTRATION NUMBER:  30,024
;     REFERENCE/DOCKET NUMBER:  WPB 28835
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703) 836-6400
;     TELEFAX:  (703) 836-2787
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  268 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single-stranded
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   ORIGINAL SOURCE:
;     ORGANISM:  Mycobacterium bovis
;     STRAIN:
;     INDIVIDUAL ISOLATE:
;   POSITION IN GENOME:
;     CHROMOSOME/SEGMENT:
;     MAP POSITION:
;   FEATURE:
;     NAME/KEY:
;     LOCATION:  461-728
;     IDENTIFICATION METHOD:
;     OTHER INFORMATION:
US-08-105-168B-1


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 268;
  Best Local Similarity   88.2%;  Pred. No. 2.3e-02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCGG 17
           ||||||||||  |||| |
Db      76 GGCCAGCACGGTGGCCG 60


Search completed: June 22, 2000, 15:22:05
Job time: 18803 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:38 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-19
Perfect score: 18
Sequence: 1 GGCCAGCACGCTGGCGGG 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	16.4	91.1	150	81	AW424994	AW424994 50914 MAR
	2	16.4	91.1	520	84	B07741	B07741 8025PI17092
	3	16	88.9	818	60	AI793994	AI793994 fc36b11.y
	4	15.4	85.6	201	25	N64111	N64111 yv60e08.r1
	5	15.4	85.6	336	50	AU069006	AU069006 AU069006
	6	15.4	85.6	373	22	R35412	R35412 yg65c09.r1
	7	15.4	85.6	416	107	AQ500061	AQ500061 HS_5220_A
	8	15.4	85.6	436	82	FR0024593	AL017437 F.rubripe
	9	15.4	85.6	451	47	AI546492	AI546492 LD48078.5
	10	15.4	85.6	457	40	AA941865	AA941865 LD27176.5
c	11	15.4	85.6	460	94	AQ005918	AQ005918 CIT-HSP-2
	12	15.4	85.6	535	107	AQ492803	AQ492803 HS_5222_B
	13	15.4	85.6	536	33	AA441316	AA441316 LD16206.5
	14	15.4	85.6	544	47	AI541803	AI541803 SD06979.5
	15	15.4	85.6	548	37	AA694692	AA694692 GM01307.5
c	16	15.4	85.6	552	80	AW355374	AW355374 pnf-b.pk0
	17	15.4	85.6	553	42	AI107513	AI107513 GH05235.5
	18	15.4	85.6	571	34	AA464442	AA464442 zx78h11.r
	19	15.4	85.6	581	44	AI260560	AI260560 LP04471.5
	20	15.4	85.6	594	41	AI062439	AI062439 GH01644.5
	21	15.4	85.6	594	82	FR0024572	AL017416 F.rubripe
	22	15.4	85.6	660	47	AI533336	AI533336 SD05110.5
	23	15.4	85.6	770	37	AA735369	AA735369 LD21194.5
c	24	15.4	85.6	873	69	AW116200	AW116200 fi12e01.x
	25	15	83.3	272	40	AA977589	AA977589 on61c03.s
	26	15	83.3	336	34	AA527988	AA527988 ne61b10.s
	27	15	83.3	416	36	AA627540	AA627540 nq50d02.s
	28	15	83.3	423	36	AA627536	AA627536 nq50c09.s
	29	15	83.3	431	40	AA928416	AA928416 on49g07.s
	30	15	83.3	457	35	AA580073	AA580073 nl79h02.s
c	31	15	83.3	480	80	AW315478	AW315478 12856 MAR
	32	15	83.3	482	36	AA644367	AA644367 af64c01.s
c	33	15	83.3	506	62	AI894177	AI894177 mc54h02.x
	34	15	83.3	590	39	AA857037	AA857037 of06g04.s
	35	14.8	82.2	124	24	H95103	H95103 yw57b12.r1
c	36	14.8	82.2	162	33	AA416754	AA416754 zu08h06.s
	37	14.8	82.2	178	60	AV130183	AV130183 AV130183
c	38	14.8	82.2	182	20	Z20058	Z20058 HSAAABIGW S
	39	14.8	82.2	205	20	T58083	T58083 yb26a04.r1
	40	14.8	82.2	209	20	X61806	X61806 MUSTSG148X
	41	14.8	82.2	211	59	AV115849	AV115849 AV115849
	42	14.8	82.2	221	59	AV115516	AV115516 AV115516
	43	14.8	82.2	226	72	AV335674	AV335674 AV335674
	44	14.8	82.2	227	38	AA776036	AA776036 ae84g10.s
	45	14.8	82.2	233	72	AV294427	AV294427 AV294427

ALIGNMENTS

```
RESULT   1
AW424994/c
LOCUS       AW424994       150 bp    mRNA            EST       10-FEB-2000
DEFINITION  50914 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW424994
VERSION     AW424994.1  GI:6952941
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A. and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138228.
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 23  row: H  column: 17
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       26 a      54 c      44 g      26 t
ORIGIN


  Query Match             91.1%;  Score 16.4;  DB 81;  Length 150;
  Best Local Similarity   94.4%;  Pred. No. 5.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCGGG 18
           |||||||||| ||||||
Db      89 GGCCAGCACGCGGGCGGG 72


RESULT   2
B07741
LOCUS       B07741         520 bp    DNA             GSS       15-NOV-1997
DEFINITION  8025PI17092895T7 Rhodobacter sphaeroides 2.4.1 genomic DNA library
            Rhodobacter sphaeroides genomic clone 8025PI1709289ST7 similar to
            dppB (P37316), genomic survey sequence.
ACCESSION   B07741
VERSION     B07741.1  GI:2055533
KEYWORDS    GSS.
SOURCE      Rhodobacter sphaeroides.
  ORGANISM  Rhodobacter sphaeroides
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodobacter.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.
            and Kaplan,S.
  TITLE     Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
            chromosome II is a true chromosome
  JOURNAL   Microbiology 143, 3085-3099 (1997)
  MEDLINE   98015398
COMMENT     Contact: Mackenzie, C.
            Department of Microbiology and Molecular Genetics
            The University of Texas Health Science Center
            6431 Fannin St., Houston, TX 77006, USA
            Tel: (713) 794-1742
            Fax: (713) 794-1782
            Email: mackenz@utmmg.med.uth.tmc.edu
            Seq primer: pBluescript SK (-) T7
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Rhodobacter sphaeroides"
                     /strain="2.4.1T"
                     /db_xref="taxon:1063"
                     /clone="8025PI17092895T7"
                     /clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
                     library"
                     /lab_host="E. coli S17-1"
                     /note="Vector: pLA2917; Genomic DNA from Rhodobacter
```

sphaeroides was prepared and partially digested with Sau3AI. Size selected (20kb) fragments were subcloned into the BglII site of the cosmid vector pLA2917 (Allen, L. N. and R. S. Hanson. 1985. Journal of Bacteriology, 161:955-962. The library was then ordered around chromosome II of Rhodobacter sphaeroides (Choudhary et al., 1994. Journal of Bacteriology, 176:7694-7702). The cosmids were then digested with a variety of restriction enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and DNaseI) and restriction fragments subcloned into the respective multiple cloning site sites of pBluescript SK (-). Note BglII fragments were subcloned into the pBluescript BamHI site. DNaseI fragments were subcloned into the EcoRV site. All subclones were transformed into E. coli XL1Blue MRF'. All fragments were then sequenced and the sequences where possible were assembled using the GCG program GELASSEMBLE. "
BASE COUNT 110 a 162 c 174 g 67 t 7 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 84; Length 520;
Best Local Similarity 94.4%; Pred. No. 7.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCCAGCACGCTGGCGGG 18
||||||||||.|||||||
Db 172 GGCCAGCACGATGGCGGG 189

RESULT 3
AI793994
LOCUS AI793994 818 bp mRNA EST 02-JUL-1999
DEFINITION fc36b11.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to TR:Q13436 Q13436 NUCLEOLAR RNA HELICASE GU ;, mRNA sequence.
ACCESSION AI793994
VERSION AI793994.1 GI:5341720
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE 1 (bases 1 to 818)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189153.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)
Possible reversed clone: similarity on wrong strand
Seq primer: T3 ET from Amersham
High quality sequence stop: 434.
FEATURES Location/Qualifiers
source 1. .818
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield stage embryos"
/lab_host="XL1-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik,Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."
BASE COUNT 191 a 213 c 236 g 148 t 30 others
ORIGIN

Query Match 88.9%; Score 16; DB 60; Length 818;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CCAGC CGCTGGCGGG 18
||||||||||||||||
Db 421 CCAGCACGCTGGCGGG 436

RESULT 4
N64111
LOCUS N64111 201 bp mRNA EST 28-JAN-1997
DEFINITION yv60e08.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:247142 5' similar to gb:M81635 ERYTHROCYTE BAND 7 INTEGRAL MEMBRANE PROTEIN (HUMAN);, mRNA sequence.
ACCESSION N64111
VERSION N64111.1 GI:1211940
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 201)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Oct 18, 1995 this sequence version replaced gi:1024175.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 695 Std Error: 0.00
Seq primer: T7

```
          High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Homo sapiens"
                     /db_xref="GDB:3796388"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:247142"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       47 a     46 c     39 g     61 t      8 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 25;  Length 201;
  Best Local Similarity   88.9%;  Pred. No. 1.7e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGGG 18
          |||||| | |||||||||
Db     89 GGCCAGNAAGCTGGCGGG 106


RESULT   5
AU069006
LOCUS       AU069006      336 bp    mRNA            EST       07-JUN-1999
DEFINITION  AU069006 Rice callus Oryza sativa cDNA clone C51634_2A, mRNA
            sequence.
ACCESSION   AU069006
VERSION     AU069006.1  GI:5003857
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from callus (1998)
  JOURNAL   Unpublished (1998)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187478.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .336
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="C51634_2A"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid. "
BASE COUNT       57 a     99 c     92 g     79 t      9 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 50;  Length 336;
  Best Local Similarity   94.1%;  Pred. No. 1.9e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          |||||||||||||| |||
Db    183 GGCCAGCACGCTGCCGG 199


RESULT   6
R35412
LOCUS       R35412        373 bp    mRNA            EST       02-MAY-1995
DEFINITION  yg65c09.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:38274 5', mRNA sequence.
ACCESSION   R35412
VERSION     R35412.1  GI:792313
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2263
            High quality sequence stops: 319 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2263   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Homo sapiens"
                     /db_xref="GDB:410815"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:38274"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       75 a     120 c     106 g     70 t      2 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 22;  Length 373;
  Best Local Similarity   94.1%;  Pred. No. 1.9e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
```

```
          ||||| ||||||||||
Db    137 GGCCATCACGCTGGCGG 153


RESULT   7
AQ500061
LOCUS       AQ500061      416 bp    DNA             GSS       28-APR-1999
DEFINITION  HS_5220_A2_E04_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=796 Col=8 Row=I, genomic survey sequence.
ACCESSION   AQ500061
VERSION     AQ500061.1  GI:4692634
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 796  row: I  column: 8
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=796 Col=8 Row=I"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      100 a    107 c     73 g    136 t
ORIGIN


  Query Match           85.6%;  Score 15.4;  DB 107;  Length 416;
  Best Local Similarity   94.1%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 GCCAGCACGCTGGCGGG 18
          |||||||| ||||||||
Db    314 GCCAGCACTCTGGCGGG 330


RESULT   8
FR0024593
LOCUS       FR0024593     436 bp    DNA             GSS       10-DEC-1997
DEFINITION  F.rubripes GSS sequence, clone 126G07cD2, genomic survey sequence.
ACCESSION   AL017437
VERSION     AL017437.1  GI:2683805
KEYWORDS    GSS; genome survey sequence.
SOURCE      Fugu rubripes.
  ORGANISM  Fugu rubripes
            Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S.J., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac:uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .436
                     /organism="Fugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 126G07"
                     /clone="126G07cD2"
BASE COUNT      127 a    111 c     98 g     93 t      7 others
ORIGIN


  Query Match           85.6%;  Score 15.4;  DB 82;  Length 436;
  Best Local Similarity   94.1%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          |||||| ||||||||||
Db    251 GGCCAGGACGCTGGCGG 267


RESULT   9
AI546492
LOCUS       AI546492      451 bp    mRNA            EST       22-MAR-1999
DEFINITION  LD48078.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD48078 5prime, mRNA sequence.
ACCESSION   AI546492
VERSION     AI546492.1  GI:4463865
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138251.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 480  row: G  column: 6
            High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD48078"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      112 a    119 c     134 g     86 t
ORIGIN
```

```
  Query Match             85.6%;  Score 15.4;  DB 47;  Length 451;
  Best Local Similarity   94.1%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          || ||||||||||||||
Db    304 GGACAGCACGCTGGCGG 320


RESULT  10
AA941865
LOCUS       AA941865      457 bp    mRNA            EST       25-NOV-1998
DEFINITION  LD27176.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD27176 5prime, mRNA sequence.
ACCESSION   AA941865
VERSION     AA941865.1  GI:3102416
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Trachata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Dec 14, 1995 this sequence version replaced gi:1122626.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 271  row: G  column: 4
            High quality sequence stop: 431.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD27176"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      112 a    121 c    136 g     88 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 40;  Length 457;
  Best Local Similarity   94.1%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          || ||||||||||||||
Db    305 GGACAGCACGCTGGCGG 321


RESULT  11
AQ005918/c
LOCUS       AQ005918      460 bp    DNA             GSS       27-JUN-1998
DEFINITION  CIT-HSP-2297O16.TR CIT-HSP Homo sapiens genomic clone 2297O16,
            genomic survey sequence.
ACCESSION   AQ005918
VERSION     AQ005918.1  GI:3083363
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2297O16.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /db_xref="GDB:7153532"
                     /db_xref="taxon:9606"
                     /clone="2297O16"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       92 a    150 c     80 g    138 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 94;  Length 460;
  Best Local Similarity   94.1%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          ||||||||||||||| ||
Db    126 GGCCAGCACGCTGGAGG 110


RESULT  12
AQ492803
LOCUS       AQ492803      535 bp    DNA             GSS       28-APR-1999
DEFINITION  HS_5222_B1_A04_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=798 Col=7 Row=B, genomic survey sequence.
ACCESSION   AQ492803
VERSION     AQ492803.1  GI:4693028
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
```

```
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 798  row: B  column: 7
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 535.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=798 Col=7 Row=B"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT       99 a    142 c    154 g    133 t      7 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 107;  Length 535;
  Best Local Similarity   94.1%;  Pred. No. 2.1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          ||||||| |||||||||
Db    312 GGCCAGCCCGCTGGCGG 328


RESULT  13
AA441316
LOCUS       AA441316      536 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD16206.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD16206 5prime, mRNA sequence.
ACCESSION   AA441316
VERSION     AA441316.1  GI:2153194
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395311.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 162  row: A  column: 6
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln015425"
                     /db_xref="taxon:7227"
                     /clone="LD16206"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT       128 a    144 c    163 g    101 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 33;  Length 536;
  Best Local Similarity   94.1%;  Pred. No. 2.1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          || |||||||||||||||
Db    304 GGACAGCACGCTGGCGG 320


RESULT  14
AI541803
LOCUS       AI541803      544 bp    mRNA            EST       22-MAR-1999
DEFINITION  SD06979.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD06979 5prime, mRNA
            sequence.
ACCESSION   AI541803
VERSION     AI541803.1  GI:4459176
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948690.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 69  row: G  column: 7
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD06979"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT       129 a    149 c    165 g    101 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 47;  Length 544;
  Best Local Similarity   94.1%;  Pred. No. 2.1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCCAGCACGCTGGCGG 17
          || |||||||||||||||
Db    304 GGACAGCACGCTGGCGG 320


RESULT  15
AA694692
LOCUS       AA694692      548 bp    mRNA            EST       28-NOV-1998
DEFINITION  GM01307.5prime GM Drosophila melanogaster ovary BlueScript
            Drosophila melanogaster cDNA clone GM01307 5prime, mRNA sequence.
ACCESSION   AA694692
VERSION     AA694692.1  GI:2697277
KEYWORDS    EST.
SOURCE      fruit fly.
```

```
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1405018.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 13  row: A  column: 7
            High quality sequence stop: 547.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM01307"
                     /clone_lib="GM Drosophila melanogaster ovary BlueScript"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="SOLR"
                     /note="Organ: ovary; Vector: BlueScript SK; Site_1: EcoRI;
                     Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      135 a    147 c    165 g    101 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DI 37;  Length 548;
  Best Local Similarity   94.1%;  Pred. No. 2.1e 03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGCCAGCACGCTGGCGG 17
           || |||||||||||||||
Db     316 GGACAGCACGCTGGCGG 332



  Search completed: June 22, 2000, 12:08:42
  Job time: 10436 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:40 ; Search time 8627.09 Seconds
(without alignments)
-2.030 Million cell updates/sec

Title: US-09-362-485-20
Perfect score: 18
Sequence: 1 CACCCGTTCGGACAGTAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87623	A87623 Sequence 20
2	c	18	100.0	1194	2	MTU92472	U92472 Mycobacteri
3	c	18	100.0	1208	5	A89752	A89752 Sequence 9
4	c	18	100.0	1209	5	A87612	A87612 Sequence 9
5	c	18	100.0	1228	5	A87608	A87608 Sequence 5
6	c	18	100.0	1228	5	A89748	A89748 Sequence 5
7	c	18	100.0	1229	5	A87610	A87610 Sequence 7
8	c	18	100.0	1229	5	A89750	A89750 Sequence 7
9	c	18	100.0	1235	5	A87606	A87606 Sequence 3
10	c	18	100.0	1235	5	A87609	A87609 Sequence 6
11	c	18	100.0	1235	5	A87611	A87611 Sequence 8
12	c	18	100.0	1235	5	A89746	A89746 Sequence 3
13	c	18	100.0	1235	5	A89749	A89749 Sequence 6
14	c	18	100.0	1235	5	A89751	A89751 Sequence 8
15	c	18	100.0	1236	5	A87613	A87613 Sequence 10
16	c	18	100.0	1236	5	A89753	A89753 Sequence 10
17	c	18	100.0	1237	5	A87607	A87607 Sequence 4
18	c	18	100.0	1237	5	A89747	A89747 Sequence 4
19	c	18	100.0	1245	5	A87605	A87605 Sequence 2
20	c	18	100.0	1245	5	A89745	A89745 Sequence 2
21	c	18	100.0	1260	5	A87604	A87604 Sequence 1
22	c	18	100.0	1260	5	A89744	A89744 Sequence 1
23	c	18	100.0	2584	1	MTALADH	X63069 M.tuberculo
24	c	18	100.0	56414	1	MTV002	AL008967 Mycobacte
25	c	15	83.3	128011	45	AC009122	AC009122 Homo sapi
26	c	14.8	82.2	438	9	HUMD4F11	D16933 Human HepG2
27		14.8	82.2	592	13	CNS01GWK	AL143685 Anopheles
28		14.8	82.2	620	13	G55749	G55749 SHGC-100957
29	c	14.8	82.2	1847	12	MUSCHAP	L25913 Mus musculu
30	c	14.8	82.2	1883	11	HSU38846	U38846 Human stimu
31	c	14.8	82.2	1966	12	MMCCTDE	Z31554 M.musculus
32	c	14.8	82.2	1986	5	A45918	A45918 Sequence 57
33	c	14.8	82.2	2003	40	AF026291	AF026291 Homo sapi
34	c	14.8	82.2	20588	2	YPS251712	AJ251712 Yersinia
35	c	14.8	82.2	20745	2	YPE251713	AJ251713 Yersinia
36	c	14.8	82.2	40121	1	MSGY414A	AD000007 Mycobacte
37		14.8	82.2	43430	1	MTCY16B7	Z81331 Mycobacteri
38	c	14.8	82.2	85561	8	ATAC009326	AC009326 Arabidops
39		14.8	82.2	131043	33	AL133458	AL133458 Homo sapi
40		14.8	82.2	154329	33	AL136972	AL135972 Homo sapi
41		14.8	82.2	177367	41	AC009706	AC009706 Homo sapi
42	c	14.8	82.2	205214	55	AC017106	AC017106 Homo sapi
43	c	14.8	82.2	216500	32	CNS01DRE	AL117192 Homo sapi
44	c	14.4	80.0	486	8	SCQRGSS34B	L05970 Sclerotium
45	c	14.4	80.0	931	13	CNS01JFY	AL146991 Anopheles

ALIGNMENTS

```
RESULT   1
A87623
LOCUS       A87623          18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 20 from Patent WO9836089.
ACCESSION   A87623
VERSION     A87623.1  GI:6736263
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      6 c      4 g      3 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db      1 CACCCGTTCGGACAGTAA 18
```

```
RESULT   2
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,. . and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108
```

```
RESULT   3
A89752/c
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108
```

```
RESULT   4
A87612/c
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match          100.0%;  Score 18;  DB 5;  Length 1209;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108


RESULT   5
A87608/c
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match          100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108


RESULT   6
A89748/c
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match          100.0%;  Score 18;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108


RESULT   7
A87610/c
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match          100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT   8
A89750/c
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match          100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT   9
A87606/c
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
```

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108


RESULT  10
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT  11
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT  12
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108


RESULT  13
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT  14
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1124 CACCCGTTCGGACAGTAA 1107


RESULT  15
A87613/c
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match            100.0%;  Score 18;  DB 5;  Length 1236;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1125 CACCCGTTCGGACAGTAA 1108
```

```
Search completed: June 22, 2000, 14:58:42
Job time: 17937 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:59 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-20
Perfect score: 18
Sequence: 1 CACCCGTTCGGACAGTAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49521	Mycobacterium sp.
2		18	100.0	18	1	V49619	AlaDH derived olig
c 3		18	100.0	1245	1	V49626	Mycobacterium tube
c 4		18	100.0	1260	1	V49510	Mycobacterium sp.
c 5		18	100.0	1260	1	V49625	Mycobacterium tube
c 6		14.8	82.2	1986	1	Q97797	Clone pTdelta2 cod
c 7		14	77.8	17955	1	V56642	Actinoplanes sp. a
c 8		13.8	76.7	972	1	T34189	PcpD coding sequen
c 9		13.8	76.7	3635	1	Q80004	Sucrose-phosphate-
c 10		13.4	74.4	360	1	Q66413	alcA promoter. DNA
c 11		13.4	74.4	644	1	N60396	Sequence of promot
c 12		13.4	74.4	1[illegible]3	1	N60399	Sequence of the pr
c 13		13.4	74.4	2280	1	N60395	Sequence constitut
14		13.4	74.4	2854	1	V32591	Schwannomin-bindin
15		13.4	74.4	2915	1	Q23261	Carrot aspartokina
16		13.4	74.4	2915	1	X00761	Carrot aspartokina
c 17		13.4	74.4	6611	1	T09311	Mycobacteriophage
c 18		13.4	74.4	6611	1	T51223	NheI-D fragment of
c 19		13.4	74.4	6611	1	T66129	Mycobacteriophage
c 20		13.4	74.4	6611	1	T70491	Mycobacteriophage
c 21		13.4	74.4	110000	1	V21209_03	Continuation (4 of
c 22		13.2	73.3	266	1	V20278	Probe (179) for mi
c 23		13.2	73.3	329	1	Q50578	Asparaginylendopep
c 24		13.2	73.3	594	1	Q05691	DNA encoding MG-4
c 25		13.2	73.3	594	1	N92571	DNA sequenc of M7
c 26		13.2	73.3	700	1	X30339	DNA encoding a hum
c 27		13.2	73.3	1910	1	Q50579	Asparaginylendopep
c 28		13.2	73.3	3025	1	Q13337	D1 dopamine recept
c 29		13.2	73.3	3025	1	T63657	D1 dopamine recept
30		13.2	73.3	6360	1	Q20577	Human Cytomegalovi
31		13.2	73.3	6360	1	T91283	Human cytomegalovi
c 32		13.2	73.3	6412	1	T28796	Human calcium sens
c 33		13.2	73.3	6412	1	V06017	Human calcium prot
c 34		13.2	73.3	19186	1	X20620	Polynucleotide seq
35		13.2	73.3	28720	1	V49655	Human SC3 DNA. Pro
c 36		13	72.2	1245	1	T29923	Saccharothrix aero
37		12.8	71.1	504	1	Q61049	Human brain Expres
38		12.8	71.1	1058	1	T91508	Mycobacterium tube
39		12.8	71.1	1058	1	T91444	Mycobacterium tube
40		12.8	71.1	1058	1	V44383	Mycobacterium tube
41		12.8	71.1	1058	1	V64491	M. tuberculosis im
c 42		12.8	71.1	1275	1	T78903	FE65 cDNA. Peptide
c 43		12.8	71.1	1275	1	V36874	Nucleotide sequenc
c 44		12.8	71.1	1500	1	T78902	Carboxy-terminal r
c 45		12.8	71.1	1500	1	V36873	Nucleotide sequenc

ALIGNMENTS

RESULT 1
V49521
ID V49521 standard; DNA; 18 BP.
AC V49521;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R3.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 5 A; 6 C; 4 G; 3 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||||||||||||||||
Db 1 CACCCGTTCGGACAGTAA 18

RESULT 2
V49619
ID V49619 standard; DNA; 18 BP.
AC V49619;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R3.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 5 A; 6 C; 4 G; 3 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||||||||||||||||
Db 1 CACCCGTTCGGACAGTAA 18

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||||||||||||||||
Db 1131 CACCCGTTCGGACAGTAA 1114

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||||||||||||||||
Db 1146 CACCCGTTCGGACAGTAA 1129

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match            100.0%;  Score 18;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 0.18;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||||||||||||||||
Db   1146 CACCCGTTCGGACAGTAA 1129


RESULT   6
Q97797/c
ID   Q97797 standard; cDNA; 1986 BP.
AC   Q97797;
DT   12-JUL-1996  (first entry)
DE   Clone pTdelta2 coding for mouse CCT-delta subunit.
KW   Chaperonin containing TCP-1; CCT; subunit; protein folding complex;
KW   murine; denaturation; renaturation; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             162. .1781
FT                   /*tag=  a
FT                   /product= CCT-delta
PN   WO9520654-A1.
PD   03-AUG-1995.
PF   31-JAN-1995; G00192.
PR   31-JAN-1994; GB-001791.
PR   09-SEP-1994; GB-018234.
PA   (CANC-) INST CANCER RES ROYAL CANCER HOSPITAL.
PI   Ashworth A,  Kubota H,  Willison KR;
DR   WPI; 95-275439/36.
DR   P-PSDB; R79580.
PT   Folding protein complex sub-unit(s) - with ability to form complex
PT   in vitro, useful in facilitating folding of proteins
PS   Claim 2; Fig 8d; 94pp; English.
CC   Mouse cDNAs encoding novel TCP-1-related proteins were isolated
CC   using a combination of methods. Clones pTdelta2 and pTzeta12 were
CC   isolated by cross-hybridisation with C.elegans cDNA probes recovered
CC   from a 5'-expressed sequence tag collection. The clones pTbeta2,
CC   pTepsilon5 and pTthetal were isolated by cross-hybridisation with
CC   human probes recovered by PCR of human HT1080 cell line cDNA using
CC   degenerate primers from a conserved region of TCP-1 and TF55. Clone
CC   pCBL80 was isolated during a mouse testis cDNA sequencing project
CC   and clone pTgamma7 was recovered by hybridisation with a mouse cDNA
CC   PCR product which was made with primers derived from the sequence of
CC   a TCP-1 related gene fragment recovered accidentally from a human
CC   kidney cDNA library during a screen for ion transport channel
CC   genes. TCP-1 and the proteins encoded by the cDNA clones are
CC   individual subunits of a heteromeric chaperonin which has been
CC   designated "chaperonin containing TCP-1" (CCT). The chaperonin acts
CC   as a protein folding complex. The present sequence is that of cDNA
CC   clone pTdelta2 which encodes the CCT-delta subunit.
SQ   Sequence   1986 BP;    582 A;    423 C;    476 G;    505 T;


  Query Match            82.2%;  Score 14.8;  DB 1;  Length 1986;
  Best Local Similarity  88.9%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          |||||  ||||||||  |||
Db   1656 CACCCTTTCGGACATTAA 1639


RESULT   7
V56642/c
ID   V56642 standard; DNA; 17955 BP.
AC   V56642;
DT   26-NOV-1998  (first entry)
DE   Actinoplanes sp. acarbose acb cluster DNA.
KW   Acarbose; acb cluster; acbQ; acbK; acbM; acbL; acbC; acbA; acbB; acbD;
KW   acbE; acbG; acbF; acbH; maltodextrin transferase; dTDP-glucose synthase;
KW   dTDP-Glucose-4,6-dehydratase; C7 sugar cyclase; acarviosyl transferase;
KW   alpha amylase; MalF-like membrane protein; MalG-like membrane protein;
KW   MalE-like binding protein; acarbose-7-kinase; oxidoreductase;
KW   maltodextrin glucotransferase; ss.
OS   Actinoplanes sp.
FH   Key             Location/Qualifiers
FT   CDS             1. .1585
FT                   /*tag=  a
FT                   /product= "acbQ
FT                   /note= "Maltodextrin glucotransferase. No start codon
FT                          given"
FT   CDS             1617. .2519
FT                   /*tag=  b
FT                   /product= "acbK"
FT                   /note= "Acarbose-7-kinase"
FT   CDS             2516. .3727
FT                   /*tag=  c
FT                   /product= "acbM"
FT                   /note= "Protein of unknown function with GTG start codon"
FT   CDS             3724. .4722
FT                   /*tag=  d
FT                   /product= "acbL"
FT                   /note= "Partial sequence of oxidoreductase"
FT   CDS             4719. .6104
FT                   /*tag=  e
FT                   /product= "acbN"
FT                   /note= "protein of unknown function as given in
FT                          specification. Coding region is interrupted
FT                          with stop codons"
FT   CDS             6101. .7249
FT                   /*tag=  f
FT                   /product= "acbC"
FT                   /note= "C7 sugar cyclase with GTG start codon"
FT   CDS             complement (7291. .8316)
FT                   /*tag=  g
FT                   /product= "acbB"
FT                   /note= "dTDP-glucose-4,6-dehydratase"
FT   CDS             8333. .9310
FT                   /*tag=  h
FT                   /product= "acbA"
FT                   /note= "dTDP-glucose synthase"
FT   CDS             complement (9455. .12520)
FT                   /*tag=  i
FT                   /product= "acbE"
FT                   /note= "alpha amylase"
FT   CDS             12846. .15020
FT                   /*tag=  j
FT                   /product= "acbD"
FT                   /note= "acarviosyl transferase with GTG start codon"
FT   CDS             complement (15255. .16013)
FT                   /*tag=  k
FT                   /product= "acbG"
FT                   /note= "MalG-like membrane protein with GTG start codon"
FT   CDS             complement (16010. .16987)
FT                   /*tag=  l
FT                   /product= "acbF"
FT                   /note= "MalF-like membrane protein"
FT   CDS             complement (16987. .17955)
FT                   /*tag=  m
FT                   /product= "acbH"
FT                   /note= "MalE-like binding protein. No start codon given"
PN   DE19708127-A1.
PD   03-SEP-1998.
PF   28-FEB-1997; 008127.
```

PR 28-FEB-1997; DE-008127.
PA (FARB) BAYER AG.
PI Apeler H, Crueger A, Distler J, Goeke K, Jarling M,
PI Pape H, Paz Marta DU, Piepersberg W, Schroeder W,
PI Stratmann A;
DR WPI; 98-468360/41.
DR P-PSDB; W76186, W76187, W76188, W76189, W76190, W76191, W76192, W76193,
DR W76194, W76195, W76196, W76197.
PT New actinoplanes acarbose gene cluster - useful for increasing
PT acarbose production in Actinoplanes, etc
PS Claim 1; Fig 3; 51pp; German.
CC This sequence is an Actinoplanes acarbose gene cluster which encodes
CC the acbA, acbB, acbC, acbD, acbE, acbF, acbG, acbH, acbK, acbL, acbM
CC and acbQ proteins. The gene cluster is useful for increasing acarbose
CC production in Actinoplanes or other microorganisms or in vitro
CC synthesis systems.
SQ Sequence 17955 BP; 2621 A; 6111 C; 6527 G; 2696 T;

Query Match 77.8%; Score 14; DB 1; Length 17955;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 CCGTTCGGACAGTA 17
||||||||||||||
Db 26 CCGTTCGGACAGTA 13

RESULT 8
T34189/c
ID T34189 standard; DNA; 972 BP.
AC T34189;
DT 25-SEP-1996 (first entry)
DE PcpD coding sequence.
KW Pentachlorophenol breakdown pathway; PcpC; PcpA; PcpB; environment;
KW tetrachloro-p-hydroquinone reductase; PCP-degrading enzyme complex;
KW Flavobacterium; PCP; 2,3,5,6-tetrachloro-p-hydroquinone; TeCH;
KW glutathione; 2,3,6-trichloro-p-hydroquinone; TrCH; food chain;
KW 2,6-dichloro-p-hydroquinone; wood preserving industry; fungicide;
KW pesticide; herbicide; disinfectant; ds.
OS Flavobacterium sp. Strain ATCC 39723.
FH Key Location/Qualifiers
FT misc_difference 754
FT /*tag= a
FT /note= "Residue not given in the specification, included
FT to maintain open reading frame and to encode the
FT amino acid given in the specification at this
FT position (Glu)"
PN US5512478-A.
PD 30-APR-1996.
PF 23-MAR-1992; 856015.
PR 23-MAR-1992; US-856015.
PR 13-JUL-1992; US-914282.
PR 18-JUL-1994; US-276887.
PA (IDAH-) IDAHO RES FOUND INC.
PI Lange CC, Orser CS, Xun L;
DR WPI; 96-229872/23.
DR P-PSDB; R99487.
PT Flavobacterium sp. penta:chloro:phenol breakdown pathway genes and
PT enzymes - useful in the bio-remediation and dechlorination of PCP
PT contg. matter
PS Disclosure; Columns 61-64; 52pp; English.
CC The sequences given in T34189-90 encode proteins from the
CC pentachlorophenol (Pcp) breakdown pathway, designated PcpD and
CC PcpR. The PcpD gene was found to code for a 323 amino acid
CC polypeptide, mol. wt. 35942 daltons. Based on multiple sequence
CC alignments, PcpD belongs to the family of class 1 dioxygenase electron
CC transport proteins. PcpD was predicted to have three structural
CC domains, one involved in binding flavin mononucleotide (FMN), the second
CC responsible for binding reduced NADPH and the third forming a plant-
CC ferredoxin-type (2Fe-2S) centre. PcpD is classified as an iron
CC sulphur flavoprotein-oxidoreductase, and has been designated the PCP
CC 4-monooxygenase reductase. It is transcribed as a dicistronic message
CC with PcpB (see also T34184). The pcpR open reading frame encodes a 303
CC amino acid protein with a mol. wt. of 33549 daltons, which possesses a
CC helix-turn-helix motif in its N-terminal portion. PcpR is thought to
CC activate the transcription of pcpBD and pcpA. In combination with pcpA,
CC pcpB and pcpC, these enzymes catalyse the initial steps of PCP breakdown.
CC These enzymes can specifically be used in the breakdown of PCP
CC containing matter which persists in the environment and in food chains
CC after its use in the wood preserving industry as a fungicide and
CC pesticide, and in products such as herbicides and disinfectants.
SQ Sequence 972 BP; 181 A; 310 C; 313 G; 168 T;

Query Match 76.7%; Score 13.8; DB 1; Length 972;
Best Local Similarity 88.2%; Pred. No. 46;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTA 17
||||||||||||| | |
Db 659 CACCCGTTCGGACGGCA 643

RESULT 9
Q80004/c
ID Q80004 standard; cDNA; 3635 BP.
AC Q80004;
DT 05-AUG-1995 (first entry)
DE Sucrose-phosphate-synthase.
KW Sucrose-phosphate-synthase; sugarbeet; sucrose; crop improvement;
KW ss.
OS Beta vulgaris.
FH Key Location/Qualifiers
FT cds 30. .3167
FT /*tag= a
PN DE4317596-A.
PD 01-DEC-1994.
PF 24-MAY-1993; 317596.
PR 24-MAY-1993; DE-317596.
PA (SCHD) SCHERING AG.
PA (AGRE) HOECHST SCHERING AGREVO GMBH.
PI Hesse H, Mueller-Roeber B;
DR WPI; 95-007485/02.
DR P-PSDB; R66221.
PT New DNA sequences encoding sucrose regulating enzymes of sugar
PT beet - for regulating sucrose concn. specifically ADP glucose
PT pyrophosphorylase, sucrose phosphate synthase and sucrose
PT synthase.
PS Claim 3; Page 64-72; 78pp; German.
CC cDNA encoding sugarbeet sucrose-phosphate-synthase was
CC isolated from a cDNA library in phage lambda Zap. Alterations
CC to the coding sequence may be designed to by-pass the plant's
CC regulatory mechanisms, allowing manipulation of the sucrose
CC content of sugarbeet.
SQ Sequence 3635 BP; 1085 A; 646 C; 851 G; 1053 T;

Query Match 76.7%; Score 13.8; DB 1; Length 3635;
Best Local Similarity 88.2%; Pred. No. 52;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ACCCGTTCGGACAGTAA 18
| |||||| ||||||||
Db 2844 ATCCGTTCTGACAGTAA 2828

RESULT 10
Q66413/c
ID Q66413 standard; DNA; 360 BP.
AC Q66413;
DT 26-JAN-1995 (first entry)
DE alcA promoter.
KW Promoter; alcA; creA binding site; glucose repression; ds.
OS Aspergillus nidulans.
PN WO9413820-A.

```
PD   23-JUN-1994.
PF   10-DEC-1993; E03553.
PR   10-DEC-1992; US-988778.
PA   (KONN ) GIST-BROCADES NV.
PI   Hintz WE,  Lagosky PA;
DR   WPI; 94-217893/26.
PT   DNA constructs comprising a filamentous fungal promoter variant
PT   having a disrupted creA binding site - mediates expression of the
PT   protein in the presence of glucose
PS   Disclosure; Page 13; 21pp; English.
CC   A region of the A. nidulans alcA promoter is given in Q66413.  To
CC   construct an alcA promoter variant, the creA binding sites were
CC   disrupted, but spatially conserved, by replacement of residues.
CC   The promoter variant allows expression of foreign genes in
CC   Aspergillus nidulans without the need to establish glucose-depleted
CC   growth conditions.
SQ   Sequence   360 BP;    97 A;    111 C;    83 G;    69 T;

  Query Match             74.4%;  Score 13.4;  DB 1;  Length 360;
  Best Local Similarity   93.3%;  Pred. No. 72;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db     164 CCCGTGCGGACAGTA 150


RESULT  11
N60396/c
ID   N60396 standard; DNA; 644 BP.
AC   N60396;
DT   25-JUN-1991  (first entry)
DE   Sequence of promoter region derived from alcA, synthetic signal
DE   peptide coding region on pALCA1SENDO ATCC 53370.
KW   Filamentous fungi promoter; ds.
FH   Key             Location/Qualifiers
FT   signal_peptide  549. .598
FT                   /*tag=  a
FT   cds             599. .639
FT                   /*tag=  b
FT   mat_peptide     640. .644
FT                   /*tag=  c
FT                   /note= "5' end"
FT   misc_feature    1
FT                   /*tag=  d
FT                   /note= "Base no.= 1381"
PN   WO8606097-A.
PD   23-OCT-1986.
PF   14-APR-1986; G00209.
PR   15-APR-1985; CA-479135.
PR   20-DEC-1985; US-811404.
PA   (ALLE-) ALLELIX INC.
PI   Gwynne DI, Buxton F, Pickett M, Davies R, Scazzocchio C;
DR   WPI; 86-291664/44.
DR   P-PSDB; P60460.
PT   DNA construct for use in filamentous fungi - comprising promoter
PT   operative in filamentous fungi to promote transcription of coding
PT   region
PS   Example; Fig 15; 75pp; English.
CC   In the constructs of the invention, the promoter region naturally
CC   associated with the alcohol dehydrogenase I (alcA) gene and the
CC   aldehyde dehydrogenase (aldA) gene of A. nidulans or naturally
CC   associated with the glucoamylase gene in Aspergillus niger may be
CC   used. The DNA construct may contain a promoter region in operative
CC   association with a signal peptide coding region. The promoter/signal
CC   construct is suitably provided with a flanking restriction site to
CC   allow precise coupling of the protein coding region to the signal
CC   peptide coding region.
SQ   Sequence   644 BP;    150 A;    196 C;    165 G;    133 T;

  Query Match             74.4%;  Score 13.4;  DB 1;  Length 644;
  Best Local Similarity   93.3%;  Pred. No. 76;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db     399 CCCGTGCGGACAGTA 385


RESULT  12
N60399/c
ID   N60399 standard; DNA; 1863 BP.
AC   N60399;
DT   25-JUN-1991  (first entry)
DE   Sequence of the promoter region 22 of the alcA gene, the synthetic
DE   signal peptide coding region 68 and the coding region 60 coding for
DE   human interferon alpha-2 in pALCA1SIFN.
KW   Filamentous fungi promoter; ds.
FH   Key             Location/Qualifiers
FT   misc_feature    29
FT                   /*tag=  a
FT                   /note= " = base 1200"
FT   signal_peptide  878. .922
FT                   /*tag=  b
FT   cds             923. .975
FT                   /*tag=  c
FT   mat_peptide     976. .1474
FT                   /*tag=  d
PN   WO8606097-A.
PD   23-OCT-1986.
PF   14-APR-1986; G00209.
PR   15-APR-1985; CA-479135.
PR   20-DEC-1985; US-811404.
PA   (ALLE-) ALLELIX INC.
PI   Gwynne DI, Buxton F, Pickett M, Davies R, Scazzocchio C;
DR   WPI; 86-291664/44.
PT   DNA construct for use in filamentous fungi - comprising promoter
PT   operative in filamentous fungi to promote transcription of coding
PT   region
PS   Disclosure; Fig 11; 75pp; English.
CC   In the constructs of the invention, the promoter region naturally
CC   associated with the alcohol dehydrogenase I (alcA) gene and the
CC   aldehyde dehydrogenase (aldA) gene of A. nidulans or naturally
CC   associated with the glucoamylase gene in Aspergillus niger may be
CC   used. The DNA construct may contain a promoter region in operative
CC   association with a signal peptide coding region. The promoter/signal
CC   construct is suitably provided with a flanking restriction site to
CC   allow precise coupling of the protein coding region to the signal
CC   peptide coding region.
SQ   Sequence   1863 BP;    516 A;    425 C;    422 G;    500 T;

  Query Match             74.4%;  Score 13.4;  DB 1;  Length 1863;
  Best Local Similarity   93.3%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db     668 CCCGTGCGGACAGTA 654


RESULT  13
N60395/c
ID   N60395 standard; DNA; 2280 BP.
AC   N60395;
DT   25-JUN-1991  (first entry)
DE   Sequence constituting the coding region and promoter region of the
DE   alcohol dehydrogenase I (alcA) gene of Aspergillus nidulans.
KW   Filamentous fungi promoter; ss.
OS   Aspergillus nidulans.
FH   Key             Location/Qualifiers
FT   promoter        502. .801
FT                   /*tag=  a
FT   tata_signal     693. .699
```

```
FT                   /*tag=  b
FT   exon            802. .1452
FT                   /*tag=  c
FT   exon            1501. .1851
FT                   /*tag=  d
FT   exon            1921. .1968
FT                   /*tag=  e
PN   WO8606097-A.
PD   23-OCT-1986.
PF   14-APR-1986; G00209.
PR   15-APR-1985; CA-479135.
PR   20-DEC-1985; US-811404.
PA   (ALLE-) ALLELIX INC.
PI   Gwynne DI, Buxton F, Pickett M, Davies R, Scazzocchio C;
DR   WPI; 86-291664/44.
DR   P-PSDB; P60455.
PT   DNA construct for use in filamentous fungi - comprising promoter
PT   operative in filamentous fungi to promote transcription of coding
PT   region
PS   Disclosure; Fig 1A; 75pp; English.
CC   In the constructs of the invention, the promoter region naturally
CC   associated with the alcohol dehydrogenase I (alcA) gene and the
CC   aldehyde dehydrogenase (aldA) gene of A. nidulans or naturally
CC   associated with the glucoamylase gene in Aspergillus niger may be
CC   used. The DNA construct may contain a promoter region in operative
CC   association with a signal peptide coding region. The promoter/signal
CC   construct is suitably provided with a flanking restriction site to
CC   allow precise coupling of the protein coding region to the signal
CC   peptide coding region.
SQ   Sequence   2280 BP;    535 A;    602 C;    600 G;    543 T;


  Query Match            74.4%;  Score 13.4;  DB 1;  Length 2280;
  Best Local Similarity  93.3%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 CCCGTTCGGACAGTA 17
          ||||| |||||||||
Db    608 CCCGTGCGGACAGTA 594


RESULT  14
V32591
ID   V32591 standard; cDNA; 2854 BP.
AC   V32591;
DT   28-SEP-1998  (first entry)
DE   Schwannomin-binding protein cDNA.
KW   Schwannomin-binding protein; SBP; human; neurofibromatosis 2;
KW   NF2; tumour suppressor protein; brain tumour; glioma; schwannoma;
KW   meningioma; sarcoma; melanoma; ependymoma; cataract; hamartoma;
KW   therapy; diagnosis; transgenic animal; signal transduction;
KW   cell proliferation; translation; ss; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             45  .2786
FT                   /*tag=  a
PN   WO9821333-A2.
PD   22-MAY-1998.
PF   14-NOV-1997; U21005.
PR   15-NOV-1996; US-030987.
PA   (CEDA-) CEDARS SINAI MEDICAL CENT.
PI   Pulst SM, Scoles DR;
DR   WPI; 98-297937/26.
DR   P-PSDB; W48928.
PT   New nucleic acid encoding schwannomin-binding protein - useful for,
PT   e.g. diagnosis and treatment of brain tumours and in drug screening
PS   Claim 3; Page 53-57; 75pp; English.
CC   This cDNA sequence encodes a human schwannomin-binding protein
CC   (SBP) (see W48928) that has been found to have functional
CC   involvement in the inititation of translation and is suggested to
CC   serve a role in the final steps of a signal transduction cascade
CC   affecting cell division and proliferation.  The cDNA was isolated
CC   from a human adult brain cDNA library using a yeast two-hybrid
CC   method with schwannomin isoform 1 as bait.  Novel SBPs of the
CC   invention (see W48927-31) bind to the neurofibromatosis 2 (NF2)
CC   tumour suppressor protein schwannomin, and have a role in cellular
CC   pathways that effect cell division and proliferation.  They provide
CC   targets for treating a broad variety of cancer pathologies, such as
CC   glioma, carcinoma, sarcoma, melanoma and hamartoma.  SBPs are used
CC   in a claimed method for modulating the activity of an oncogenic
CC   protein.  SBP nucleic acids, oligonucleotides (including antisense),
CC   vectors containing these, transformed host cells and SBP polypeptides
CC   and antibodies of the invention can be used to screen for potental
CC   antagonists and agonists of SBP.  These can be used to treat brain
CC   tumours, such as glioma, schwannoma, meningioma and ependymoma, as
CC   well as cataracts and retinal disorders.  Oligonucleotides that
CC   specifically hybridise with nucleic acids of the invention (see
CC   V32590-94) are used to identify or quantify nucleic acids encoding
CC   SBP, e.g. in library screening, and may also be used to detect SBP
CC   genes mutations, particularly for diagnosis of cancer.  Antisense
CC   nucleic acids are used to inhibit SBP expression.  Single-stranded
CC   primers based on SBP nucleic acids are also claimed.  Transgenic
CC   animals expressing exogenous nucleic acid encoding a SBP are models
CC   for studying activity of SBP and for drug screening.
SQ   Sequence   2854 BP;    767 A;    756 C;    807 G;    524 T;


  Query Match            74.4%;  Score 13.4;  DB 1;  Length 2854;
  Best Local Similarity  93.3%;  Pred. No. 86;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          |||| ||||||||||
Db     62 CACCGGTTCGGACAG 76


RESULT  15
Q23261
ID   Q23261 standard; cDNA; 2915 BP.
AC   Q23261;
DT   05-AUG-1992  (first entry)
DE   Carrot aspartokinase-homoserine dehydrogenase.
KW   AK-HSDH; amino acid; metabolism; biosynthesis; EC.1.1.1.3;
KW   EC.2.7.2.4.; ss.
OS   Daucus carota L. cv Danvers.
FH   Key             Location/Qualifiers
FT   cds             2. .2593
FT                   /*tag=  a
FT                   /function= EK-HSDH
PN   US7746705-A.
PD   18-FEB-1992.
PF   18-FEB-1992; 180292.
PR   16-AUG-1991; US-746705.
PA   (USDA ) US AGRIC RES SERV.
PI   Matthews B, Weisemann J;
DR   WPI; 92-113940/14.
DR   P-PSDB; R22357.
PT   Purified bi:functional protein from carrots - with aspartokinase
PT   and homoserine dehydrogenase activities
PS   Example 7; Fig 8; 57pp; English.
CC   Carrot cell culture- and carrot root cell-derived cDNA was PCR
CC   amplified. The two primers (see Q23262 and Q23263) were based on
CC   the amino acid sequences of directly sequenced peptide fragments
CC   obtained from purified HSDH. A 1100bp product was obtained and was
CC   used as a probe to screen a carrot cDNA library. A 2915bp sequence
CC   was determined from overlapping positive clones and an amino acid
CC   sequence was deduced from the nucleotide sequence.
SQ   Sequence   2915 BP;    830 A;    527 C;    704 G;    854 T;


  Query Match            74.4%;  Score 13.4;  DB 1;  Length 2915;
  Best Local Similarity  93.3%;  Pred. No. 86;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          |||||||||| ||||
```

Db 2423 CACCCGTTCGCACAG 2437

Search completed: June 22, 2000, 15:13:02
Job time: 18108 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:22:05 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-20
Perfect score: 18
Sequence: 1 CACCCGTTCGGACAGTAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	13.8	76.7	972	1	US-08-276-887A-12	Sequence 12, Appl
c	2	13.8	76.7	3635	3	US-08-553-436A-5	Sequence 5, Appli
c	3	13.4	74.4	60	1	US-08-470-958-12	Sequence 12, Appl
c	4	13.4	74.4	360	2	US-08-321-474-10	Sequence 10, Appl
c	5	13.4	74.4	1182	7	5198345-23	Patent No. 5198345
c	6	13.4	74.4	1662	7	5198345-20	Patent No. 5198345
	7	13.4	74.4	2915	1	US-07-746-705A-16	Sequence 16, Appl
	8	13.4	74.4	2915	3	US-08-380-182-18	Sequence 18, Appl
c	9	13.4	74.4	3377	7	5198345-16	Patent No. 5198345
c	10	13.4	74.4	6611	1	US-08-402-282-2	Sequence 2, Appli
c	11	13.4	74.4	6611	1	US-08-508-004-2	Sequence 2, Appli
c	12	13.4	74.4	6611	1	US-08-402-066-2	Sequence 2, Appli
c	13	13.4	74.4	6611	1	US-08-402-068-2	Sequence 2, Appli
c	14	13.2	73.3	3025	1	US-08-444-734A-1	Sequence 1, Appli
c	15	13	72.2	1245	3	US-08-750-524-2	Sequence 2, Appli
	16	12.8	71.1	1789	2	US-08-455-543A-29	Sequence 29, Appl
	17	12.8	71.1	1789	3	US-08-223-305C-29	Sequence 29, Appl
c	18	12.8	71.1	1924	3	US-08-756-317-1	Sequence 1, Appli
c	19	12.8	71.1	2181	7	5208144-36	Patent No. 5208144
	20	12.8	71.1	2338	2	US-08-455-543A-31	Sequence 31, Appl
	21	12.8	71.1	2338	3	US-08-223-305C-31	Sequence 31, Appl
	22	12.8	71.1	5904	1	US-07-745-206A-6	Sequence 6, Appli
	23	12.8	71.1	5904	2	US-08-455-543A-3	Sequence 3, Appli
	24	12.8	71.1	5904	3	US-08-193-078B-3	Sequence 3, Appli
	25	12.8	71.1	5904	3	US-08-223-305C-3	Sequence 3, Appli
	26	12.8	71.1	5904	3	US-08-149-097D-3	Sequence 3, Appli
	27	12.8	71.1	5904	3	US-08-311-363-6	Sequence 6, Appli
	28	12.4	68.9	375	2	US-08-482-882-65	Sequence 65, Appl
	29	12.4	68.9	375	3	US-08-483-389-65	Sequence 65, Appl
	30	12.4	68.9	375	3	US-08-487-113D-65	Sequence 65, Appl
	31	12.4	68.9	375	3	US-08-473-503-65	Sequence 65, Appl
	32	12.4	68.9	375	3	US-08-483-932-65	Sequence 65, Appl
	33	12.4	68.9	375	4	US-08-720-420A-65	Sequence 65, Appl
	34	12.4	68.9	375	5	US-08-714-017-65	Sequence 65, Appl
	35	12.4	68.9	731	7	5304637-3	Patent No. 5304637
	36	12.4	68.9	1333	2	US-08-910-973-21	Sequence 21, Appl
c	37	12.4	68.9	1680	4	US-09-052-339-2	Sequence 2, Appli
c	38	12.4	68.9	1826	4	US-09-052-339-3	Sequence 3, Appli
	39	12.4	68.9	8855	3	US-08-542-003-1	Sequence 1, Appli
	40	12.4	68.9	8855	3	US-08-322-760A-1	Sequence 1, Appli
c	41	12.2	67.8	339	3	US-08-428-197-45	Sequence 45, Appl
c	42	12.2	67.8	339	6	PCT-US93-10555-45	Sequence 45, Appl
c	43	12.2	67.8	362	1	US-08-053-131-170	Sequence 170, App
c	44	12.2	67.8	362		US-08-096-762-170	Sequence 170, App
c	45	12.2	67.8	381	3	US-08-053-131-155	Sequence 155, App

ALIGNMENTS

RESULT 1
US-08-276-887A-12/c
; Sequence 12, Application US/08276887A
; Patent No. 5512478
; GENERAL INFORMATION:
; APPLICANT: Orser, Cindy S. and Xun, Luying
; TITLE OF INVENTION: GENES AND ENZYMES INVOLVED
; TITLE OF INVENTION: IN THE MICROBIAL
; TITLE OF INVENTION: DEGRADATION OF
; TITLE OF INVENTION: PENTACHLOROPHENOL
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ramon A. Klitzke II
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1/PC Gene
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,887A
; FILING DATE: July 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/914,282
; FILING DATE: July 13, 1992
; APPLICATION NUMBER: 07/856,015
; FILING DATE: March 23, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramon A. Klitzke II
; REGISTRATION NUMBER: 30,188
; REFERENCE/DOCKET NUMBER: 2815-36746
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 972 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double-stranded
; TOPOLOGY: Linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: No

```
;    ANTI-SENSE:  No
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  Flavobacterium sp. Strain
;      ORGANISM:  ATCC 39723
US-08-276-887A-12


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 972;
  Best Local Similarity   88.2%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CACCCGTTCGGACAGTA 17
           ||||||||||||| | |
Db     659 CACCCGTTCGGACGGCA 643


RESULT   2
US-08-553-436A-5/c
; Sequence 5, Application US/08553436A
; Patent No. 5866790
;  GENERAL INFORMATION:
;    APPLICANT:  HESSE, Holger
;    APPLICANT:  MULLER-ROBER, Bernd
;    TITLE OF INVENTION:  DNA SEQUENCES AND PLASMIDS FOR THE
;    TITLE OF INVENTION:  PREPARATION OF SUGAR BEET WITH CHANGED SUCROSE
;    TITLE OF INVENTION:  CONCENTRATION
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/553,436A
;      FILING DATE:  17-NOV-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/01671
;      FILING DATE:  20-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 4317596.1
;      FILING DATE:  24-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-117
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3635 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Beta vulgaris
;      INDIVIDUAL ISOLATE:  Saccharosephosphate Synthase
;    IMMEDIATE SOURCE:
;      LIBRARY:  phage lamda zap
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  30..3167
US-08-553-436A-5


  Query Match             76.7%;  Score 13.8;  DB 3;  Length 3635;
  Best Local Similarity   88.2%;  Pred. No. 30;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 ACCCGTTCGGACAGTAA 18
           | ||||||| ||||||||
Db    2844 ATCCGTTCTGACAGTAA 2828


RESULT   3
US-08-470-958-12/c
; Sequence 12, Application US/08470958
; Patent No. 5674707
;  GENERAL INFORMATION:
;    APPLICANT:  HINTZ, WILLIAM E.
;    APPLICANT:  LAGOSKY, PETER A.
;    TITLE OF INVENTION:  PRODUCTION OF HETEROLOGOUS PROTEINS IN
;    TITLE OF INVENTION:  FILAMENTOUS FUNGI
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE
;      STREET:  8TH FLOOR, 1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,958
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MITCHARD, LEONARD C.
;      REGISTRATION NUMBER:  29,009
;      REFERENCE/DOCKET NUMBER:  1459-6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NXN UR
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  60 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-470-958-12


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 60;
  Best Local Similarity   93.3%;  Pred. No. 31;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db      44 CCCGTGCGGACAGTA 30


RESULT   4
US-08-321-474-10/c
; Sequence 10, Application US/08321474
; Patent No. 5710021
;  GENERAL INFORMATION:
;    APPLICANT:  Hintz, William E.
;    APPLICANT:  Lagosky, Peter A.
;    TITLE OF INVENTION:  PRODUCTION OF HETEROLOGOUS PROTEINS IN
```

```
;    TITLE OF INVENTION:  FILAMENTOUS FUNGI
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye
;      STREET:  8th Floor, 1100 No. 5710021th Glebe Road
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,474
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/988,778
;      FILING DATE:  10-DEC-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C.
;      REGISTRATION NUMBER:  29009
;      REFERENCE/DOCKET NUMBER:  617-13
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  360 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-321-474-10


  Query Match           74.4%;  Score 13.4;  DB 2;  Length 360;
  Best Local Similarity  93.3%;  Pred. No. 38;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db     164 CCCGTGCGGACAGTA 150


RESULT   5
5198345-23/c
;Patent No. 5198345
;    APPLICANT: GWYNNE, DAVID I.;BUXTON, FRANCIS P.;PICKETT, MARK H.
;DAVIES, ROGER W.;SCAZZOCCHIO, CLAUDIO
;    TITLE OF INVENTION: VECTORS IN USE IN FILAMENTOUS FUNGI
;    NUMBER OF SEQUENCES: 28
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/811,404
;      FILING DATE: 20-DEC-1985
;SEQ ID NO:23:
;      LENGTH: 1182
5198345-23


  Query Match           74.4%;  Score 13.4;  DB 7;  Length 1182;
  Best Local Similarity  93.3%;  Pred. No. 44;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db     503 CCCGTGCGGACAGTA 489


RESULT   6
5198345-20/c
;Patent No. 5198345
;    APPLICANT: GWYNNE, DAVID I.;BUXTON, FRANCIS P.;PICKETT, MARK H.
;DAVIES, ROGER W.;SCAZZOCCHIO, CLAUDIO
;    TITLE OF INVENTION: VECTORS IN USE IN FILAMENTOUS FUNGI
;    NUMBER OF SEQUENCES: 28
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/811,404
;      FILING DATE: 20-DEC-1985
;SEQ ID NO:20:
;      LENGTH: 1662
5198345-20


  Query Match           74.4%;  Score 13.4;  DB 7;  Length 1662;
  Best Local Similarity  93.3%;  Pred. No. 46;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CCCGTTCGGACAGTA 17
           ||||| |||||||||
Db    1178 CCCGTGCGGACAGTA 1164


RESULT   7
US-07-746-705A-16
; Sequence 16, Application US/07746705A
; Patent No. 5451516
;  GENERAL INFORMATION:
;    APPLICANT:  Matthews, Benjamin F.
;    APPLICANT:  Weisemann, Jane M.
;    TITLE OF INVENTION:  A Recombinant DNA Molecule Encoding
;    TITLE OF INVENTION:  a Bifunctional Plant Enzyme: Aspartokinase and Homoserine
;    TITLE OF INVENTION:  Dehydrogenase
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janelle S. Graeter
;      STREET:  Bldg. 005. Roon 402, BARC-W
;      CITY:  Beltsville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20705
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/746,705A
;      FILING DATE:  19910816
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Graeter, Janelle S.
;      REGISTRATION NUMBER:  35,024
;      REFERENCE/DOCKET NUMBER:  4000.91
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301)504-5676
;      TELEFAX:  (301)504-5060
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2915 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..2593
US-07-746-705A-16


  Query Match           74.4%;  Score 13.4;  DB 1;  Length 2915;
```

```
  Best Local Similarity   93.3%;  Pred. No. 49;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          ||||||||||  ||||
Db   2423 CACCCGTTCGCACAG 2437


RESULT   8
US-08-380-182-18
; Sequence 18, Application US/08380182
; Patent No. 5858749
;  GENERAL INFORMATION:
;    APPLICANT:  Matthews, Benjamin F.
;    APPLICANT:  Weisemann, Jane M.
;    TITLE OF INVENTION:  A Bifunctional Protein From Carrots
;    TITLE OF INVENTION:  (Daucus carota) with Aspartokinase and Homoserine
;    TITLE OF INVENTION:  Dehydrogenase Activities"
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Janelle S. Graeter
;      STREET:  Room 411,BGldg. 005, BARC-W
;      CITY:  Beltsville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20705
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/380,182
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Graeter, Janelle S.
;      REGISTRATION NUMBER:  35,024
;      REFERENCE/DOCKET NUMBER:  0226.94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-504-6629
;      TELEFAX:  301-504-5060
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2915 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Daucus carota
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..2593
US-08-380-182-18


  Query Match            74.4%;  Score 13.4;  DB 3;  Length 2915;
  Best Local Similarity   93.3%;  Pred. No. 49;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          ||||||||||  ||||
Db   2423 CACCCGTTCGCACAG 2437


RESULT   9
5198345-16/c
;Patent No. 5198345
;    APPLICANT: GWYNNE, DAVID I.;BUXTON, FRANCIS P.;PICKETT, MARK H.
;DAVIES, ROGER W.;SCAZZOCCHIO, CLAUDIO
;    TITLE OF INVENTION: VECTORS IN USE IN FILAMENTOUS FUNGI
;    NUMBER OF SEQUENCES: 28
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/811,404
;      FILING DATE: 20-DEC-1985
;SEQ ID NO:16:
;      LENGTH: 3377
5198345-16


  Query Match            74.4%;  Score 13.4;  DB 7;  Length 3377;
  Best Local Similarity   93.3%;  Pred. No. 50;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 CCCGTTCGGACAGTA 17
          |||||  |||||||||
Db   1176 CCCGTGCGGACAGTA 1162


RESULT   10
US-08-402-282-2/c
; Sequence 2, Application US/08402282
; Patent No. 5476768
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,282
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6611 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  390..1538
;      OTHER INFORMATION:  /function= "coding sequence"
;      OTHER INFORMATION:  /product= "DNA polymerase"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2107..3132
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3138..3359
```

```
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4690..5028
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5028..5375
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5375..5653
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5653..5910
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6078..6491
;       OTHER INFORMATION:  /function= "potential open reading
;       OTHER INFORMATION:  frame"
US-08-402-282-2


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 6611;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          ||||||||||| |||
Db   2655 CACCCGTTCGGCCAG 2641


RESULT  11
US-08-508-004-2/c
; Sequence 2, Application US/08508004
; Patent No. 5582969
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/508,004
;      FILING DATE:  27-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/402,282
;      FILING DATE:  10-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6611 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  390..1538
;      OTHER INFORMATION:  /function= "coding sequence"
;      OTHER INFORMATION:  /product= "DNA polymerase"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2107..3132
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3138..3359
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4690..5028
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5028..5375
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5375..5653
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5653..5910
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6078..6491
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-508-004-2


  Query Match             74.4%;  Score 13.4;  DB 1;  Length 6611;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          ||||||||||| |||
Db   2655 CACCCGTTCGGCCAG 2641


RESULT  12
US-08-402-066-2/c
; Sequence 2, Application US/08402066
; Patent No. 5612182
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
```

```
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,066
;      FILING DATE:
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6611 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  390..1538
;      OTHER INFORMATION:  /function= "coding sequence"
;      OTHER INFORMATION:  /product= "DNA polymerase"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2107..3132
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3138..3359
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4690..5028
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5028..5375
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5375..5653
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5653..5910
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6078..6491
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-402-066-2

  Query Match             74.4%;  Score 13.4;  DB 1;  Length 6611;
  Best Local Similarity   93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CACCCGTTCGGACAG 15
           ||||||||||| |||
Db    2655 CACCCGTTCGGCCAG 2641


RESULT  13
US-08-402-068-2/c
; Sequence 2, Application US/08402068
; Patent No. 5633159
;  GENERAL INFORMATION:
;    APPLICANT:  Pearson, Robert E.
;    APPLICANT:  Dickson, Julie A.
;    APPLICANT:  Hamilton, Paul T.
;    APPLICANT:  Little, Michael C.
;    APPLICANT:  Beyer Jr., Wayne F.
;    TITLE OF INVENTION:  MYCOBACTERIOPHAGE SPECIFIC FOR THE
;    TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS COMPLEX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  ,135
;      REFERENCE/DOCKET NUMBER:  P-3283
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6611 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  390..1538
;      OTHER INFORMATION:  /function= "coding sequence"
;      OTHER INFORMATION:  /product= "DNA polymerase"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2107..3132
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3138..3359
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4690..5028
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5028..5375
```

```
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5375..5653
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5653..5910
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6078..6491
;      OTHER INFORMATION:  /function= "potential open reading
;      OTHER INFORMATION:  frame"
US-08-402-068-2


  Query Match            74.4%;  Score 13.4;  DB 1;  Length 6611;
  Best Local Similarity  93.3%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAG 15
          ||||||||||| |||
Db   2655 CACCCGTTCGGCCAG 2641


RESULT  14
US-08-444-734A-1/c
; Sequence 1, Application US/08444734A
; Patent No. 5610282
;  GENERAL INFORMATION:
;    APPLICANT:  Sibley, David R.
;    APPLICANT:  Monsma, Frederick J.
;    APPLICANT:  Mahan, Lawrence C.
;    APPLICANT:  McVittie, Loris D.
;    TITLE OF INVENTION:  cDNA encoding the rat D1 dopamine
;    TITLE OF INVENTION:  receptor linked to adnylyl cyclase activation and
;    TITLE OF INVENTION:  expression of the receptor protein in plasmid-transfected
;    TITLE OF INVENTION:  cell lines
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson and Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,734A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/029,917
;      FILING DATE:  03-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/548,714
;      FILING DATE:  06-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Altman, Daniel E.
;      REGISTRATION NUMBER:  34,115
;      REFERENCE/DOCKET NUMBER:  NIH065.001FW1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3025 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  pB73D1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  293..1756
US-08-444-734A-1


  Query Match            73.3%;  Score 13.2;  DB 1;  Length 3025;
  Best Local Similarity  83.3%;  Pred. No. 64;
  Matches   15;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          |||| ||| |||||| ||
Db   2072 CACCTGTTTGGACAGCAA 2055


RESULT  15
US-08-750-524-2/c
; Sequence 2, Application US/08750524
; Patent No. 5861293
;  GENERAL INFORMATION:
;    APPLICANT:  Katsuhisa KOJIRI, et al.
;    TITLE OF INVENTION:  GENE ENCODING GLYCOSYLTRANSFERASE AND ITS
;    TITLE OF INVENTION:  USES
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Office of Sherman and Shalloway
;      STREET:  413 N. Washington Street
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22314
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  Dell System 210; Intel 80 286 Microprocessor
;      OPERATING SYSTEM:  MS DOS 6.2
;      SOFTWARE:  Word Perfect, Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/750,524
;      FILING DATE:  December 12, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard A. Steinberg
;      REGISTRATION NUMBER:  26,588
;      REFERENCE/DOCKET NUMBER:  S-2401
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 549-2282
;      TELEFAX:  (703) 836-0106
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1245 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharothrix aerocolonigenes
;    FEATURE:
;      NAME/KEY:  ATCC39243
US-08-750-524-2


  Query Match            72.2%;  Score 13;  DB 3;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3 CCCGTTCGGACAG 15
           !!!!!!!!!!!!!
Db    1132 CCCGTTCGGACAG 1120
```

Search completed: June 22, 2000, 15:22:08
Job time: 18806 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:42 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-20
Perfect score: 18
Sequence: 1 CACCCGTTCGGACAGTAA 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	15	83.3	398	74	AV390744	AV390744 AV390744
	2	15	83.3	569	74	AV391799	AV391799 AV391799
	3	14.8	82.2	217	45	AI346865	AI346865 qp53f11.x
	4	14.8	82.2	238	36	AA610007	AA610007 af18c06.s
	5	14.8	82.2	241	34	AA470740	AA470740 ne13h01.s
	6	14.8	82.2	268	37	AA668379	AA668379 ab84a01.s
	7	14.8	82.2	268	41	AI039974	AI039974 ox88e11.x
c	8	14.8	82.2	338	30	AA198281	AA198281 mv49a08.r
c	9	14.8	82.2	340	22	R33195	R33195 yh77f07.r1
	10	14.8	82.2	348	39	AA833719	AA833719 aj48a11.s
c	11	14.8	82.2	352	32	AA353761	AA353761 EST61979
c	12	14.8	82.2	354	32	AA363853	AA363853 EST74511
	13	14.8	82.2	354	63	AW011736	AW011736 ul90g06.x
c	14	14.8	82.2	356	102	AQ103635	AQ103635 HS_3093_A
c	15	14.8	82.2	360	26	W74781	W74781 zd57b07.r1
	16	14.8	82.2	368	41	AU018393	AU018393 AU018393
	17	14.8	82.2	376	61	AI843565	AI843565 UI-M-AO1-
	18	14.8	82.2	378	26	W20406	W20406 zc68b07.s1
	19	14.8	82.2	380	79	AW254603	AW254603 UI-R-BJ0-
c	20	14.8	82.2	384	37	AA691547	AA691547 vs16c03.r
	21	14.8	82.2	387	28	AA112387	AA112387 zn68f12.s
	22	14.8	82.2	396	63	AI952390	AI952390 wx73b06.x
c	23	14.8	82.2	397	34	AA513816	AA513816 ne84g02.s
	24	14.8	82.2	402	42	AI146006	AI146006 UI-R-BT0-
c	25	14.8	82.2	404	36	AA638927	AA638927 vn76c02.r
	26	14.8	82.2	407	48	AI554496	AI554496 tq22c03.x
c	27	14.8	82.2	409	22	R26002	R26002 yh44c12.r1
	28	14.8	82.2	411	26	W32432	W32432 zc05d12.s1
	29	14.8	82.2	411	70	AW152420	AW152420 xg64c05.x
	30	14.8	82.2	412	48	AI559174	AI559174 tq42e08.x
	31	14.8	82.2	415	28	AA083854	AA083854 zn16c08.s
c	32	14.8	82.2	417	30	AA268049	AA268049 va40f02.r
	33	14.8	82.2	420	29	AA136129	AA136129 zn88c05.s
	34	14.8	82.2	421	61	AI826691	AI826691 wk46c12.x
	35	14.8	82.2	423	24	H98735	H98735 yx13h04.s1
	36	14.8	82.2	423	26	W93065	W93065 zh45a04.s1
c	37	14.8	82.2	426	29	AA126359	AA126359 zl23d04.r
	38	14.8	82.2	427	64	AW089723	AW089723 xd11d01.x
c	39	14.8	82.2	428	26	W45007	W45007 zc05d12.r1
	40	14.8	82.2	430	39	AA889217	AA889217 ak18f11.s
	41	14.8	82.2	431	37	AA670227	AA670227 ae94g06.s
	42	14.8	82.2	432	29	AA171730	AA171730 zo95e06.s
	43	14.8	82.2	432	36	C78737	C78737 C78737 Mous
	44	14.8	82.2	434	25	N70039	N70039 za70b01.s1
	45	14.8	82.2	434	62	AI925427	AI925427 wo30b06.x

ALIGNMENTS

RESULT 1
AV390744
LOCUS AV390744 398 bp mRNA EST 21-JAN-2000
DEFINITION AV390744 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii cDNA clone CM060d09_r, mRNA sequence.
ACCESSION AV390744
VERSION AV390744.1 GI:6544960
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 398)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags
JOURNAL DNA Res. 6, 369-373 (1999)
COMMENT On May 1, 1997 this sequence version replaced gi:2059638.
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .398
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM060d09_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 99 a 102 c 121 g 76 t
ORIGIN

Query Match 83.3%; Score 15; DB 74; Length 398;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACCCGTTCGGACAGT 16
||||||||||||||||
Db 148 ACCCGTTCGGACAGT 162

RESULT 2
AV391799
LOCUS AV391799 569 bp mRNA EST 21-JAN-2000
DEFINITION AV391799 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii cDNA clone CM078a02_r, mRNA sequence.
ACCESSION AV391799
VERSION AV391799.1 GI:6546015
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 569)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags
JOURNAL DNA Res. 6, 369-373 (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3137811.
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .569
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM078a02_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 128 a 156 c 180 g 105 t
ORIGIN

Query Match 83.3%; Score 15; DB 74; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACCCGTTCGGACAGT 16
||||||||||||||||
Db 142 ACCCGTTCGGACAGT 156

```
RESULT   3
AI346865
LOCUS       AI346865      217 bp    mRNA            EST       02-FEB-1999
DEFINITION  qp53f11.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1926765 3'
            similar to SW:TCPD_HUMAN P50991 T-COMPLEX PROTEIN 1, DELTA SUBUNIT
            ;, mRNA sequence.
ACCESSION   AI346865
VERSION     AI346865.1  GI:4084071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2287162.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 1767   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .217
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1926765"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     colon adenocarcinoma, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       54 a     52 c     49 g     62 t
ORIGIN


  Query Match           82.2%;  Score 14.8;  DB 45;  Length 217;
  Best Local Similarity  88.9%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||| |||||||| |||
Db    121 CACCCTTTCGGACATTAA 138


RESULT   4
AA610007
LOCUS       AA610007      238 bp    mRNA            EST       02-MAR-1998
DEFINITION  af18c06.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1032010
            3' similar to SW:TCPD_MOUSE P80315 T-COMPLEX PROTEIN 1, DELTA
            SUBUNIT ;, mRNA sequence.
ACCESSION   AA610007
VERSION     AA610007.1  GI:2458435
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393595.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Insert Length: 691   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1032010"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TJTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),, digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       64 a     53 c     54 g     67 t
ORIGIN


  Query Match           82.2%;  Score 14.8;  DB 36;  Length 238;
  Best Local Similarity  88.9%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||| |||||||| |||
Db    121 CACCCTTTCGGACATTAA 138


RESULT   5
AA470740
LOCUS       AA470740      241 bp    mRNA            EST       14-AUG-1997
DEFINITION  nel3h01.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:881137 3'
            similar to SW:TCPD_MOUSE P80315 T-COMPLEX PROTEIN 1, DELTA SUBUNIT
            ;, mRNA sequence.
ACCESSION   AA470740
VERSION     AA470740.1  GI:2198049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 241)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394984.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 915 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:881137"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. "

BASE COUNT 60 a 60 c 53 g 68 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 34; Length 241;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||| |||||||| |||
Db 121 CACCCTTTCGGACATTAA 138

RESULT 6
AA668379
LOCUS AA668379 268 bp mRNA EST 20-NOV-1997
DEFINITION ab84a01.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:853608 3' similar to SW:TCPD_MOUSE P80315 T-COMPLEX PROTEIN 1, DELTA SUBUNIT ;, mRNA sequence.
ACCESSION AA668379
VERSION AA668379.1 GI:2629878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Nov 6, 1997 this sequence version replaced gi:315819.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:853608"
/clone_lib="Stratagene fetal retina 937202"
/sex="mixed"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"

BASE COUNT 73 a 64 c 58 g 73 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 37; Length 268;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||| |||||||| |||
Db 121 CACCCTTTCGGACATTAA 138

RESULT 7
AI039974
LOCUS AI039974 268 bp mRNA EST 24-SEP-1998
DEFINITION ox88e11.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1663436 3' similar to SW:TCPD_HUMAN P50991 T-COMPLEX PROTEIN 1, DELTA SUBUNIT ;, mRNA sequence.
ACCESSION AI039974
VERSION AI039974.1 GI:3279168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900186.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 842 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1663436"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 74 a 64 c 56 g 73 t 1 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 41; Length 268;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||| |||||||| |||
Db 121 CACCCTTTCGGACATTAA 138

RESULT 8
AA198281/c
LOCUS AA198281 338 bp mRNA EST 22-JAN-1997
DEFINITION mv49a08.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone IMAGE:658358 5' similar to gb:L25913 Mus musculus chaperonin mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION AA198281
VERSION AA198281.1 GI:1793948
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 338)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., M rtin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394831.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:404206
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 325.
FEATURES Location/Qualifiers
source 1. .338
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:658358"
/clone_lib="Soares mouse 3NME12 5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 99 a 54 c 79 g 106 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 30; Length 338;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||| |||||||| |||
Db 24 CACCCTTTCGGACATTAA 7

RESULT 9
R33195/c
LOCUS R33195 340 bp mRNA EST 28-APR-1995
DEFINITION yh77f07.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:135781 5' similar to SP:TCPD_MOUSE P80315 T-COMPLEX PROTEIN 1, DELTA SUBUNIT ;, mRNA sequence.
ACCESSION R33195
VERSION R33195.1 GI:789053
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 340)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 635
High quality sequence stops: 297
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 635 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 297.
FEATURES Location/Qualifiers
source 1. .340
/organism="Homo sapiens"
/db_xref="GDB:541755"
/db_xref="taxon:9606"
/clone="IMAGE:135781"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 96 a 71 c 83 g 90 t
ORIGIN

```
  Query Match            82.2%;  Score 14.8;  DB 22;  Length 340;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||| |||||||| |||
Db    213 CACCCTTTCGGACATTAA 196


RESULT  10
AA833719
LOCUS       AA833719      348 bp    mRNA            EST       31-DEC-1998
DEFINITION  aj48a11.s1 Soares_testis_NHT Homo sapiens cDNA clone 1393532 3',
            mRNA sequence.
ACCESSION   AA833719
VERSION     AA833719.1  GI:2907447
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407506.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 431   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1393532"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      111 a     76 c     57 g    104 t
ORIGIN


  Query Match            82.2%;  Score 14.8;  DB 39;  Length 348;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||| |||||||| |||
Db    328 CACCCTTTCGGACATTAA 345


RESULT  11
AA353761/c
LOCUS       AA353761      352 bp    mRNA            EST       21-APR-1997
DEFINITION  EST61979 Activated T-cells XX Homo sapiens cDNA 5' end similar to
            similar to T-complex protein 1, mRNA sequence.
ACCESSION   AA353761
VERSION     AA353761.1  GI:2006121
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., ,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140210
COMMENT     On May 5, 1995 this sequence version replaced gi:797848.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):154030"
                     /db_xref="taxon:9606"
                     /clone_lib="Activated T-cells XX"
                     /cell_type="T-lymphocyte"
                     /dev_stage="adult"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       93 a     67 c     85 g     98 t      9 others
ORIGIN


  Query Match            82.2%;  Score 14.8;  DB 32;  Length 352;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CACCCGTTCGGACAGTAA 18
          ||||| |||||||| |||
Db     79 CACCCTTTCGGACATTAA 62


RESULT  12
AA363853/c
LOCUS       AA363853      354 bp    mRNA            EST       21-APR-1997
DEFINITION  EST74511 Pineal gland II Homo sapiens cDNA 5' end similar to
            similar to T-complex protein 1, mRNA sequence.
```

```
ACCESSION   AA363853
VERSION     AA363853.1  GI:2016328
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., ,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   12140200
COMMENT     On May 5, 1995 this sequence version replaced gi:797906.
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .354
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):168702"
                      /db_xref="taxon:9606"
                      /clone_lib="Pineal gland II"
                      /dev_stage="adult"
                      /note="Organ: pineal body; Vector: pBluescript SK-;
                      Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       93 a     70 c     93 g     97 t      1 others
ORIGIN
```

```
  Query Match            82.2%;  Score 14.8;  DB 32;  Length 354;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CACCCGTTCGGACAGTAA 18
           ||||| |||||||| |||
Db     325 CACCCTTTCGGACATTAA 308
```

```
RESULT  13
AW011736
LOCUS       AW011736      354 bp    mRNA            EST       10-SEP-1999
DEFINITION  ul90g06.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2159386 3' similar to gb:Z31554 M.musculus (MOUSE);, mRNA
            sequence.
ACCESSION   AW011736
VERSION     AW011736.1  GI:5860514
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133600.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1002558
            Seq primer: custom primer used.
FEATURES              Location/Qualifiers
     source           1. .354
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:2159386"
                      /clone_lib="Sugano mouse kidney mkia"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                      (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                      was primed with an oligo(dT) primer
                      [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                      ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                      and cloned into distinct DraIII sites of the pME18S-FL3
                      vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                      be used to isolate the cDNA insert.  Size selection was
                      performed to exclude fragments <1.5kb.  Library
                      constructed by Dr. Sumio Sugano (University of Tokyo
                      Institute of Medical Science).  Custom primers  for
                      sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                      primer CGACCTGCAGCTCGAGCACA."
BASE COUNT       101 a     85 c     62 g     106 t
ORIGIN
```

```
  Query Match            82.2%;  Score 14.8;  DB 63;  Length 354;
  Best Local Similarity  88.9%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CACCCGTTCGGACAGTAA 18
           ||||| |||||||| |||
Db     281 CACCCTTTCGGACATTAA 298
```

```
RESULT  14
AQ103635/c
LOCUS       AQ103635      356 bp    DNA             GSS       27-AUG-1998
DEFINITION  HS_3093_A2_D01_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3093 Col=2 Row=G, genomic survey
            sequence.
ACCESSION   AQ103635
VERSION     AQ103635.1  GI:3474664
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
```

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3093 row: G column: 2
Class: BAC ends
High quality sequence stop: 356.

FEATURES Location/Qualifiers
source 1. .356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3093 Col=2 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 129 a 62 c 108 g 57 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 102; Length 356;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||||| | ||||||||
Db 236 CACCCGTCCAGACAGTAA 219

RESULT 15
W74781/c
LOCUS W74781 360 bp mRNA EST 16-OCT-1996
DEFINITION zd57b07.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:344725 5' similar to SW:TCPD_MOUSE P80315 T-COMPLEX PROTEIN 1, DELTA SUBUNIT ;, mRNA sequence.
ACCESSION W74781
VERSION W74781.1 GI:1385013
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Nov 29, 1993 this sequence version replaced gi:502162.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 582 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 286.

FEATURES Location/Qualifiers
source 1. .360
/organism="Homo sapiens"
/db_xref="GDB:1270100"
/db_xref="taxon:9606"
/clone="IMAGE:344725"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."

BASE COUNT 101 a 77 c 87 g 93 t 2 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 26; Length 360;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CACCCGTTCGGACAGTAA 18
||||| |||||||| |||
Db 218 CACCCTTTCGGACATTAA 201

Search completed: June 22, 2000, 12:08:46
Job time: 10440 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:51 ; Search time 783.2 Seconds
(without alignments)
-22.357 Million cell updates/sec

Title: US-09-362-485-21
Perfect score: 18
Sequence: 1 CGCGGCCGACATCATCGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pri:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87624	A87624 Sequence 21
2	c	18	100.0	1245	5	A87605	A87605 Sequence 2
3	c	18	100.0	1245	5	A89745	A89745 Sequence 2
4	c	18	100.0	1260	5	A87604	A87604 Sequence 1
5	c	18	100.0	1260	5	A89744	A89744 Sequence 1
6	c	18	100.0	2584	1	MTALADH	X63069 M.tuberculo
7	c	18	100.0	24055	1	MTV036	AL021931 Mycobacte
8	c	18	100.0	56414	1	MTV002	AL008967 Mycobacte
9	c	16.4	91.1	9000	1	SC9A10	AL031260 Streptomy
10		16	88.9	778	2	MSU26405	U26405 Mycobacteri
11	c	16	88.9	1237	5	A87607	A87607 Sequence 4
12	c	16	88.9	1237	5	A89747	A89747 Sequence 4
13		16	88.9	3157	2	MSGPOLA	L11920 Mycobacteri
14	c	16	88.9	4189	1	AFANREA	L31491 Plasmid pTO
15		16	88.9	35938	1	MTCY01B2	Z95554 Mycobacteri
16		16	88.9	46423	34	CEC47A10	Z81484 Caenorhabdi
17		15.4	85.6	239	16	SVU27917	U27917 Stealth vir
18		15.4	85.6	689	16	AF065740	AF065740 Stealth v
19		15.4	85.6	1732	12	RNU52095	U52095 Rattus norv
20		15.4	85.6	1835	10	HSU17278	U17278 Human colla
21		15.4	85.6	2813	12	RNU52102	U52102 Rattus norv
22		15.4	85.6	2842	9	D78012	D78012 Homo sapien
23		15.4	85.6	3507	35	AF141673	AF141673 Drosophil
24	c	15.4	85.6	4500	1	MSP242630	AJ242630 Methyloba
25		15.4	85.6	4688	1	SYCCPN60	M57517 Synechocyst
26	c	15.4	85.6	4960	2	BLU76614	U76614 Bifdiobacte
27		15.4	85.6	7298	2	AF131877	AF131877 Streptomy
28		15.4	85.6	8420	16	SVU27802	U27802 Stealth vir
29	c	15.4	85.6	10403	34	CEK04H8	Z81563 Caenorhabdi
30		15.4	85.6	11364	1	MTV013	AL021309 Mycobacte
31		15.4	85.6	13170	1	PSECOBGEN	M62866 P.denitrifi
32		15.4	85.6	14006	2	AE002049	AE002049 Deinococc
33		15.4	85.6	28690	34	CEW08G11	Z92817 Caenorhabdi
34	c	15.4	85.6	29355	35	AC005931	AC005931 Leishmani
35	c	15.4	85.6	41284	35	AC005930	AC005930 Leishmani
36		15.4	85.6	42231	2	AF088897	AF088897 Zymomonas
37		15.4	85.6	45857	34	DMC95B7	AL021728 Drosophil
38		15.4	85.6	64411	32	CEY63D3_3	Continuation (4 of
39		15.4	85.6	77428	43	AC017634	AC017634 Drosophil
40	c	15.4	85.6	105795	1	SYCCPNC	D64001 Synechocyst
41	c	15.4	85.6	110000	32	CEY105E8_1	Continuation (2 of
42		15.4	85.6	200000	33	AC005141	AC005141 Homo sapi
43		15.4	85.6	200000	41	AC005289	AC005289 Homo sapi
44		15	83.3	20	5	A87625	A87625 Sequence 22
45		15	83.3	1216	8	AF052424	AF052424 Datisca g

ALIGNMENTS

```
RESULT   1
A87624
LOCUS       A87624         18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 21 from Patent WO9836089.
ACCESSION   A87624
VERSION     A87624.1  GI:6736264
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      8 c      5 g      2 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATCGC 18
          ||||||||||||||||||
Db      1 CGCGGCCGACATCATCGC 18


RESULT   2
A87605/c
LOCUS       A87605       1245 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9836089.
ACCESSION   A87605
VERSION     A87605.1  GI:6736245
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATCGC 18
          ||||||||||||||||||
Db   1245 CGCGGCCGACATCATCGC 1228


RESULT   3
A89745/c
LOCUS       A89745       1245 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9832862.
ACCESSION   A89745
VERSION     A89745.1  GI:6738279
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATCGC 18
          ||||||||||||||||||
Db   1245 CGCGGCCGACATCATCGC 1228


RESULT   4
A87604/c
LOCUS       A87604       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9836089.
ACCESSION   A87604
VERSION     A87604.1  GI:6736244
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Sin  ,M.
  TITLE     TEST KIT FOR TUB CULOSIS DIAGNOSIS OR TH  LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATCGC 18
          ||||||||||||||||||
Db   1260 CGCGGCCGACATCATCGC 1243


RESULT   5
A89744/c
LOCUS       A89744       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9832862.
ACCESSION   A89744
VERSION     A89744.1  GI:6738278
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           ||||||||||||||||||
Db    1260 CGCGGCCGACATCATCGC 1243


RESULT   6
MTALADH/c
LOCUS       MTALADH      2584 bp    DNA             BCT       01-DEC-1993
DEFINITION  M.tuberculosis gene for L-alanine dehydrogenase.
ACCESSION   X63069 S36765
VERSION     X63069.1  GI:44565
KEYWORDS    alanine dehydrogenase; extracellular; intracellular.
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 2584)
  AUTHORS   Andersen,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1991) A.B. Andersen, Statens Seruminstitut,
            Artillerivej 5, DK-2300 Copenhagen S, DENMARK
REFERENCE   2  (bases 1 to 2584)
  AUTHORS   Andersen,A.B., Andersen,P. and Ljungqvist,L.
  TITLE     Structure and function of a 40,000-molecular-weight protein antigen
            of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 60 (6), 2317-2323 (1992)
  MEDLINE   92267644
FEATURES             Location/Qualifiers
     source          1. .2584
                     /organism="Mycobacterium tuberculosis"
                     /strain="Erdman"
                     /isolate="TMC 107, lot 9A-2"
                     /db_xref="taxon:1773"
                     /clone_lib="lambda gt11"
                     /clone="lambda AA67"
     CDS             76. .1197
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="alanine  dehydrogenase"
                     /protein_id="CAA44791.1"
                     /db_xref="GI:44566"
                     /db_xref="SWISS-PROT:P30234"
                     /translation="MRVGIPTETKNNEFQFRVAITPAGVAELTRRGHEVLIQAGAGEG
                     SAITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAAS
                     RACTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGV
                     LMGGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIH
                     TRYSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGC
                     FEGSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAA
                     CRSNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT       463 a    871 c    819 g    431 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 1;  Length 2584;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           ||||||||||||||||||
Db    1260 CGCGGCCGACATCATCGC 1243


RESULT   7
MTV036/c
LOCUS       MTV036       24055 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 19/162.
ACCESSION   AL021931 AL123456
VERSION     AL021931.1  GI:3261526
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 24055)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 24055)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2909486.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .24055
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .>24055
                     /note="fragment designated v036. Does not represent a
                     physical clone"
     source          <1. .9
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y13E10"
     gene            complement(25. .618)
                     /gene="Rv0366c"
     CDS             complement(25. .618)
                     /gene="Rv0366c"
                     /note="Rv0366c, (MTV036.01c), len: 197. Unknown, weak
                     similarity to Haemophilus influenzae  hypothetical
```

protein HI1395. Contains PS00017 ATP/GTP-binding site motif A (P-loop) and PS00850 Glycine radical signature. FASTA scores: |P44173|YD95_HAEIN HYPOTHETICAL PROTEIN HI1395.(140 aa) opt: 152 z-score:198.5 E(): 0.0015. 27.0% identity in 126 aa overlap. TBparse score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0366c"
/protein_id="CAA17372.1"
/db_xref="GI:2909487"
/db_xref="SPTREMBL:O53701"
/translation="MKRLDLVAGPNGAGKSTFVALTLAPLLPGIVFVNADEIAKQRWP
DDPTSHAYQAAQVAADTRARLIDLGRPFIAETVFSHPSKLELIRTARTAGYTVVLHVL
VIPEGLAVERVRHRVAAGGHDVPETKIRERHRRLAELVAQAITLADGATVYDNSRLAG
PRIVAQFSGGGIIGRACWPSWTPPPLMSRWSNRPETA"

misc_feature complement(328. .354)
/gene="Rv0366c"
/note="PS00850 Glycine radical signature"

misc_feature complement(571. .594)
/gene="Rv0366c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene complement(647. .1036)
/gene="Rv0367c"

CDS complement(647. .1036)
/gene="Rv0367c"
/note="Rv0367c, (MTV036.02c), len: 129. Unknown, TBparse score is 0.850"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0367c"
/protein_id="CAA17373.1"
/db_xref="GI:2909488"
/db_xref="SPTREMBL:O53702"
/translation="MPKAVDRVTRVAADLVDSAAAEGARQSRSAKQQLDHWARVGRAV
SNQHTASRRRVEAALAGHLPMTDLTLEEGVVFNAEISAAIEERLSRTNYGDVLAAQGI
TTVALNDAGDIVEHRPDGTSVVLAATP"

RBS complement(1044. .1048)
/note="possible RBS for Rv0367c"

gene complement(1117. .2328)
/gene="Rv0368c"

CDS complement(1117. .2328)
/gene="Rv0368c"
/note="Rv0368c, (MTV036.03c), len: 403. Unknown but some similarity to gp|AJ224684|BJAJ4684_4 Bradyrhizobium japonicum cooxS, (422 aa), FASTA scores: opt: 341 z-score: 370.2 E(): 4.3e-13; 27.4% identity in 387 aa overlap. Also to M. tuberculosis hypothetical protein MTCY428_22. Contains PS00213 Lipocalin signature. 30.7% identity in 238 aa overlap. TBparse score is 0.909"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0368c"
/protein_id="CAA17374.1"
/db_xref="GI:2909489"
/db_xref="SPTREMBL:O53703"
/translation="MATPALLPGVDLAAFAAALAARLRDAGIPVSASGQASLVQALQQ
LVPRTPAALYWGARLTLVSRVDELATFDAVFASLFGVFGSAEPDGANRPPPPIAGPRT
PVAGVGHRAKRRSCAAQAQNLPWDTRSLTMASAGQGGPSRTLPDVLPSRIVARADEPF
DQFDPDDLRLLGAWLEATMARWPRRRSMRFESSPHGKRIDLRATMNASRSTGWESVLL
ARIRPRRRPRRVLLLCDVSRSMQPYAAIYLCLMRAAVLRRAGGHPEVFAFGTSLTRLT
SVLSHRSAEMALHRANARVTDRYGGTFIGRSVAALLAPPHGNALRGAVVIIASDGWDS
DPPDVLVHALTRVRRRAELLVWLNPRAAHPEFQPRAGSMAAALPYCDLFLPAHSLAGL
HQLLLALAGAR"

misc_feature complement(1798. .1839)
/gene="Rv0368c"
/note="PS00213 Lipocalin signature"

gene complement(2334. .2849)
/gene="Rv0369c"

CDS complement(2334. .2849)
/gene="Rv0369c"
/note="Rv0369c, (MTV036.04c), len: 171. Similar to ORF 4 of the Pseudomonas thermocarboxydovorans cut A, cut B and cut C gene cluster. Also see MTV036.05, MTV036.08, MTV036.09, and MTV026.10. FASTA scores gp|X77931|PTC2CUTAC_4(171 aa) opt: 226 z-score: 273.5 E(): 9.8e-08. 31.3% identity in 166 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0369c"
/protein_id="CAA17375.1"
/db_xref="GI:2909490"
/db_xref="SPTREMBL:O53704"
/translation="MPGAQLIGHEGDEYLGKVKVKVGPVTSEFSGKVHFVEQDRNQHR
AVFDAKGKEARGTGNAAATVAAQLHEVGERTRVTVDTDLKIVGKLAQFGSGMLQQVSE
KLLGQFVDSLEAELAAQSSESPQGTPPATEAAPIDLLQLADGGQLKKYGSALLAALTV
LLLIWVLRRRR"

gene complement(2950. .3846)
/gene="Rv0370c"

CDS complement(2950. .3846)
/gene="Rv0370c"
/note="Rv0370c, (MTV036.05c), len: 298. Similar to ORF 4 of the O. carboxidovorans coxMSL gene cluster (295 aa), FASTA scores, gp|X82447|OCCOXMSL4_4. opt: 851 z-score: 1035.6 E(): 0. 48.2% identity in 282 aa overlap; Also similar to gp|AJ224684|BJAJ4684_3 Bradyrhizobium japonicum cooxS, (302aa) opt: 881 z-score: 1076.4 E(): 0; 47.6% identity in 290 aa overlap; contains PS00017 ATP/GTP-binding site motif A. Highly similar to MTCY428_21 M. tuberculosis cosmid. TBparse score is 0.921"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0370c"
/protein_id="CAA17376.1"
/db_xref="GI:2909491"
/db_xref="SPTREMBL:O53705"
/translation="MTFASPDDVIRRFDEQNYLLDTGTASAIYLAVTLGRPLLLEGEP
GVGKTTAAKTLAVVLDTTLIRLQCYEGLTANEALYDWNYQRQLLSIRLAEARGKGISD
ISEADLYTEAYLVDRPILRCVRHRGPTPPVLLIDEIDRADDEFEALLLEFLGESAVTV
PELGTFLAECPPIAVLTSNRSRDLHDALRRRCLYHWIDYPGPDRAAAIVRRTVPGATA
PLIENATQFVCTARDLDLDKPPGVAETIDWVAALVALGVADLTAADSSPALASLGALA
KTPDDRTQIRDAYQAFTECSHA"

misc_feature complement(3700. .3723)
/gene="Rv0370c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene complement(3843. .4436)
/gene="Rv0371c"

CDS complement(3843. .4436)
/gene="Rv0371c"
/note="Rv0371c, (MTV036.06c), len: 197. Possible membrane protein, with weak similarity to yurE gene in Bacillus subtilis. FASTA scores: gp|Z99120|BSUB0017_134 B. subtilis complete genome(197 aa) opt: 217 z-score: 284.3 E(): 2.5e-08. 27.0% identity in 174 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0371c"
/protein_id="CAA17377.1"
/db_xref="GI:2909492"
/db_xref="SPTREMBL:O53706"
/translation="MTATQITGVVLAAGRSNRLGTPKQLLPYRDTTVLGATLDVARQA

Query Match 100.0%; Score 18; DB 1; Length 24055;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
||||||||||||||||||
Db 9156 CGCGGCCGACATCATCGC 9139

RESULT 8
MTV002/c
LOCUS MTV002 56414 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
ACCESSION AL008967 AL123456
VERSION AL008967.1 GI:3261491

KEYWORDS .
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 56414)
AUTHORS Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REMARK Erratum:[[published erratum appears in Nature 1998 Nov
12;396(6707):190]]
REFERENCE 2 (bases 1 to 56414)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT On Jun 27, 1998 this sequence version replaced gi:2624256.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES Location/Qualifiers
source 1. .56414
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
misc_feature <1. .>56414
/note="fragment designated v002. Does not represent a
physical clone"
source <1. .233
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y154"
gene complement(3. .527)
/gene="recX"
CDS complement(3. .527)
/gene="recX"
/note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar
to eg. RECX_PSEAE P37860 regulatory protein recx from
Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161
z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
overlap. Overlaps and extends CDS from ovelapping cosmid
MTCY154.16c"
/codon_start=1
/transl_table=11
/product="recX"
/protein_id="CAA15532.1"
/db_xref="GI:2624257"
/db_xref="SPTREMBL:O33280"
/translation="MTVSCPPPSTSEREEQARALCLRLLTARSRTRAELAGQLAKRGY
PEDIGNRVLDRLAAVGLVDDTDFAEQWVQSRRANAAKSKRALAAELHAKGVDDDVITT
VLGGIDAGAERGRAEKLVRARLRREVLIDDGTDEARVSRRLVAMLARRGYGQTLACEV
VIAELAAERERRRV"
gene complement(493. .2865)
/gene="recA"
CDS complement(493. .2865)
/gene="recA"
/note="Rv2737c, (MTV002.02c), len: 790 aa. recA,
identicalto RECA_MYCTU P26345 reca protein (790 aa).
Contains self-splicing protein element (intein) from 2294
to 969 (c), similar to intein II from TR:E332317
(EMBL:Y13030) DNA-directedDNA polymerase (EC 2.7.7.7) from
Thermococcus sp. (1829 aa), fasta scores; opt: 81 z-score:
235.2 E(): 6e-05, 24.6%identity in 183 aa overlap.
Contains PS00017 ATP/GTP-bindingsite motif A (P-loop),
PS00321 recA signature, and PS00881Protein splicing
signature. See Davis et al, (1992) Cell71(2):201-210"
/codon_start=1
/transl_table=11
/product="recA"
/protein_id="CAA15533.1"
/db_xref="GI:2624258"
/db_xref="SWISS-PROT:P26345"
/translation="MTQTPDREKALELAVAQIEKSYGKGSVMRLGDEARQPISVIPTG
SIALDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGVAAFIDAEHALDP
DYAKKLGVDTDSLLVSQPDTGEQALEIADMLIRSGALDIVVIDSVAALVPRAELEGEM
GDSHVGLQARLMSQALRKMTGALNNSGTTAIFINQLRDKIGVMFGSPETTTGGKALKF
YASVRMDVRRVETLKDGTNAVGNRTRVKVVKNKCLAEGTRIFDPVTGTTHRIEDVVDG
RKPIHVVAAAKDGTLHARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRA
AGELRKGDRVAQPRRFDGFGDSAPIPADHARLLGYLIGDGRDGWVGGKTPINFINVQR
ALIDDVTRIAATLGCAAHPQGRISLAIAHRPGERNGVADLCQQAGIYGKLAWEKTIPN
WFFEPDIAADIVGNLLFGLFESDGWVSREQTGALRVGYTTTSEQLAHQIHWLLLRFGV
GSTVRDYDPTQKRPSIVNGRRIQSKRQVFEVRISGMDNVTAFAESVPMWGPRGAALIQ
AIPEATQGRRRGSQATYLAAEMTDAVLNYLDERGVTAQEAAAMIGVASGDPRGGMKQV
LGASRLRRDRVQALADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTL
VAEGVVVHNCSPPFKQAEFDILYGKGISREGSLTDMGVDQGLIRKSGAWFTYEGEQLG
QGKENARNFLVENADVADEIEKKIKEKLGIGAVVTDDPSNDGVLPAPVDF"
misc_feature complement(790. .2112)
/gene="recA"
/note="recA intein"
misc_feature complement(790. .807)
/gene="recA"
/note="PS00881 Protein splicing signature"
misc_feature complement(2197. .2223)
/gene="recA"
/note="PS00321 recA signature"
misc_feature complement(2644. .2667)
/gene="recA"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
RBS complement(2873. .2877)
/note="possible RBS upstream of Rv2737c"
gene complement(3247. .3453)
/gene="Rv2738c"
CDS complement(3247. .3453)
/gene="Rv2738c"
/note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown,
N-terminus is highly similar to the N-terminus of the
upstream ORFMTV002.07c (78.4% identity in 37 aa overlap);
also similarto AL020958|SC4H8_5 Streptomyces coelicolor
cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;
39.7% identityin 63 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2738c"
/protein_id="CAA15534.1"
/db_xref="GI:2624260"
/db_xref="SPTREMBL:O33281"
/translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHVLTGFDGRSAAQAI
EDGVEPRDVWRALCADFDVPHDRW"

```
     gene            complement(3464. .4630)
                     /gene="Rv2739c"
     CDS             complement(3464. .4630)
                     /gene="Rv2739c"
                     /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable
                     transferase, similar to eg. TR:Q51560 (EMBL:L28170)
                     Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
                     fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
                     identity in 425 aaoverlap. Equivalent to Mycobacterium
                     leprae protein MLCB33.02c (392 aa); fasta scores
                     gp|Z94723|MLCB33_2 Mycobacteriumleprae cosmid B33 opt:
                     2112 z-score: 2364.5 E(): 0; 80.9%identity in 388 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2739c"
                     /protein_id="CAA15535.1"
                     /db_xref="GI:2624261"
                     /db_xref="SPTREMBL:O33282"
                     /translation="MRVAVVAGPDPGHSFPAIALCQRFRAAADTPTLFTGVEWLEAAR
                     AAGIDAVELDGLAATDRDLDAGARIHRRAAQMAVLNVPRLRALEPELVVSDVITACGG
                     MAAELLGIPWVELNPHPLYLPSKGLPPIGSGLAAGTGIRGRLRDATMRALTGRSWRAG
                     LRQRAAVRVEIGLPARDPGPLRRLIATLPALEVPRPDWPAEAVVVGPLHFEPTDRVLA
                     IPAGTGPVVVVAPSTALTGTAGLTEVALQSLTPGETVPSGSRLVVSRLSGADLTVPPW
                     AVAGLGSQAELLTRADLVICGGGHGMVAKTLLAGVPMVVVPGGGDQWEIANRVVRQGS
                     AVLIRPLTADALVAAVNEVLSSPRFREAARRAAASVAGAADPVRVCHDALALAG"
     gene            4674. .5123
                     /gene="Rv2740"
     CDS             4674. .5123
                     /gene="Rv2740"
                     /note="Rv2740, (MTV002.05), len: 149 aa. unknown, similar
                     to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
                     MLCB33.03 (B2235_F3_140) (178 aa), fasta scores; opt: 498
                     z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2740"
                     /protein_id="CAA15536.1"
                     /db_xref="GI:2624262"
                     /db_xref="SPTREMBL:O33283"
                     /translation="MAELTETSPETPETTEAIRAVEAFLNALQNEDFDTVDAALGDDL
                     VYENVGFSRIRGGRRTATLLRRMQGRVGFEVKIHRIGADGAAVLTERTDALIIGPLRV
                     QFWVCGVFEVDDGRITLWRDYFDVYDMFKGLLRGLVALVVPSLKATL"
     gene            5355. .6932
                     /gene="PE_PGRS"
     CDS             5355. .6932
                     /gene="PE_PGRS"
                     /note="Rv2741, (MTV002.06), len: 525 aa; Member of M.
                     tuberculosis PE_ PGRS subfamily, similar to many eg.
```

Query Match 100.0%; Score 18; DB 1; Length 56414;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCGGCCGACATCATCGC 18
           ||||||||||||||||||
Db   39439 CGCGGCCGACATCATCGC 39422
```

```
RESULT   9
SC9A10/c
LOCUS       SC9A10       9000 bp    DNA             BCT       11-AUG-1998
DEFINITION  Streptomyces coelicolor cosmid 9A10.
ACCESSION   AL031260
VERSION     AL031260.1  GI:3413819
KEYWORDS    16S rRNA gene; dapA; dapB; dihydrodipicolinate reductase;
            dihydrodipicolinate synthase; gpsI, guanosine pentaphosphate
            synthetase; polyribonucleotide nucleotidyltransferase; protease.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 9000)
  AUTHORS   Harris,D. and Taylor,K.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9000)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 9000)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 9A10 lies
            between 3C3 and 7C7 on the AseI-B genomic restriction fragment.
FEATURES                Location/Qualifiers
     source             1. .9000
                        /organism="Streptomyces coelicolor A3(2)"
                        /strain="A3(2)"
                        /db_xref="taxon:100226"
                        /clone="cosmid 9A10"
     gene               1. .175
                        /gene="gpsI"
     CDS                <1. .175
                        /gene="gpsI"
                        /note="SC9A10.01, gpsI, guanosine pentaphosphate
                        synthetase/ polyribonucleotide nucleotidyltransferase,
                        partial CDS, len: >57 aa; almost identical to Streptomyces
                        antibioticus gpsI TR:Q53597 (EMBL:U19858) guanosine
                        pentaphosphate synthetase (740 aa), 87.9% identity in 58
                        aa overlap. Overlaps and extends SC3C3.23."
                        /codon_start=2
                        /transl_table=11
                        /product="guanosine pentaphosphate synthetase/
                        polyribonucleotide nucleotidyltransferase"
                        /protein_id="CAA20288.1"
                        /db_xref="GI:3413820"
                        /db_xref="SPTREMBL:O86834"
```

```
                /translation="IRKLAGGKRVENVEDVLGVGAKVQVEIAEIDSRGKLSLIPVIEG
                EAGDDDKKDDADK"
     misc_feature    1. .104
                /gene="gpsI"
                /note="nominal overlap with 3C3 from: 31279 to: 31382"
     RBS             155. .159
                /gene="gpsI"
                /note="possible RBS upstream of SC9A10.02"
     gene            172. .1551
                /gene="SC9A10.02"
     CDS             172. .1551
                /gene="SC9A10.02"
                /note="SC9A10.02, probable protease, len: 459 aa; similar
                to members of the insulinase family e.g. YMXG_BACSU
                hypothetical processing protease (409 aa), fasta scores;
                opt: 648 z-score: 1449.2 E(): 0, 39.7% identity in 413 aa
                overlap and MPP2_YEAST mitochondrial processing peptidase
                beta subunit (462 aa), fasta scores; opt: 583 z-score:
                864.6 E(): 0, 29.7% identity in 448 aa overlap.
                Alternative start possible at aa 36. Contains PS00143
                Insulinase family, zinc-binding region signature and Pfam
                match to entry PF00675 Insulinase, Insulinase (proteinase
                M16), score 171.30, E-value 1.6e-47"
                /codon_start=1
                /transl_table=11
                /product="putative protease"
                /protein_id="CAA20289.1"
                /db_xref="GI:3413821"
                /db_xref="SPTREMBL:O86835"
                /translation="MTSRSATATARTSSEARAVARTQTLIKGEHGIGTVRRTTLPGGL
                RIVTETLPSVRSATFGIWAHVGSRDETPALNGATHYLEHLLFKGTRKRSALDISSAID
                AVGGEMNAFTAKEYTCYYARVLDTDLPLAIDVVCDMLTGSLIQEEDVDVERGAILEEI
                AMTEDDPGDCVHDLFAHTMFGDNALGRPVLGTVDTVNALTADRIRRFYRKHYDPTHLV
                VAAAGNVDHNKVVRQVRAAFEKSGALKDPAAQPLAPRAGRRTVRAAGRVELIGRKTEQ
                AHVILGMPGLARTDERRWAMGVLNTALGGGMSSRLFQEVREKRGLAYSVYSYTSGFAD
                CGLFGVYAGCRPSQVHDVLKICRDELDHVAEHGLTDDEIGRAVGQLQGSTVLGLEDTG
                ALMNRIGKSELCWGEQMSVDDMLARIASVTPDDVRAVARDVLGRRPSLSVIGPLKDKQ
                ASRLHDAVA"
     stem_loop       192. .231
                /gene="SC9A10.02"
                /note="hairpin loop with 18 /19 bp stem"
     misc_feature    301. .750
                /gene="SC9A10.02"
                /note="Pfam match to entry PF00675 Insulinase, Insulinase
                (proteinase M16), score 171.30, E-value 1.6e-47"
     misc_feature    367. .438
                /gene="SC9A10.02"
                /note="PS00143 Insulinase family, zinc-binding region
                signature"
     RBS             1562. .1566
                /note="possible RBS upstream of dapB"
     gene            1577. .2329
                /gene="dapB"
     CDS             1577. .2329
                /gene="dapB"
                /note="SC9A10.03, dapB, dihydrodipicolinate reductase,
                len: 250 aa; highly similar to many e.g. DAPB_CORGL
                dihydrodipicolinate reductase (EC 1.3.1.26) (248 aa),
                fasta scores; opt: 946 z-score: 1333.7 E(): 0  56.1%
                identity in 246 aa overlap. Contains PS01298
                Dihydrodipicolinate reductase signature"
                /codon_start=1
                /transl_table=11
                /product="putative dihydrodipicolinate reductase"
                /protein_id="CAA20290.1"
                /db_xref="GI:3413822"
                /db_xref="SPTREMBL:O86836"
                /translation="MSKLRVAVLGAKGRIGSEAVRAVEAAEDMELVAALGRGDGLEAL
                AESGAQVAVELTTPASVMDNLDYCLRHGIHAVVGTTGWTDERLARLNAWLDASPGTGV
                LIAPNFSIGAVLTMKFAQIAAPYFESVEVVELHHPNKVDAPSGTATRTAQLIAQARQK
                AGSAPAPDATATALDGARGANVDGVPVHAVRLRGLLAHQEVLLGAEGETLTVRHDSLH
                HSSFMPGILLGARRVVTTPGLTFGLEHFLDLN"
     misc_feature    1964. .2017
                /gene="dapB"
                /note="PS01298 Dihydrodipicolinate reductase signature"
     gene            2343. .2798
                /gene="SC9A10.04"
     CDS             2343. .2798
                /gene="SC9A10.04"
                /note="SC9A10.04, possible membrane protein, len: 151 aa;
                similar to hypothetical proteins from Mycobacterium bovis
                and tuberculosis that follow dapB, e.g. YDAB_MYCBO
                hypothetical 19.3 kd protein in dapb 3'region (177 aa),
                fasta scores; opt: 377 z-score: 696.1 E(): 1.7e-31, 47.1%
                identity in 136 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="putative membrane protein"
                /protein_id="CAA20291.1"
                /db_xref="GI:3413823"
                /db_xref="SPTREMBL:O86837"
                /translation="MRAKITYLVTAAVLVFYFLLVGSRGVLLIKTGTLLTVTFGVAVL
                ILPVIGLWFLWKNTQFVRRANQLAAELDAEGGLPVDELKRTPSGRVDRESADAVFALR
                RAETEDAPGDWRSWFRLAVAYHDARDTPRARKAMQRAIALHDGKHVEAA"
     stem_loop       2807. .2864
                /note="hairpin loop with 29 bp stem"
     misc_feature    2827. .2832
                /note="single clone coverage only"
     gene            complement(2874. .3428)
                /gene="SC9A10.05c"
     CDS             complement(2874. .3428)
                /gene="SC9A10.05c"
                /note="SC9A10.05c, probable membrane protein, len: 184 aa"
                /codon_start=1
                /transl_table=11
                /product="putative membrane protein"
                /protein_id="CAA20292.1"
                /db_xref="GI:3413824"
                /db_xref "SPTREMBL:O86838"
                /translation="MPLPFLTADRAFDTADDVALPYDDRDLWRRPYRPGPWRVGMAAL
                MLLLASFVLFAAVIIAATGGLSSAGVVFGIAVVVILCALRMLRMGAWVSAHGVRQVGF
                FTTRTVAWQQVASLRTVQQPVRWLGLPRTVQGQALLLVRQGRSAEPLSPLLTSHNSDF
                LARDGAFNRAADAVEAWADEYRRG"
     RBS             complement(3435. .3438)
                /note="possible RBS upstream of SC9A10.05c"
     gene            complement(3545. .3784)

  Query Match           91.1%;  Score 16.4;  DB 1;  Length 9000;
  Best Local Similarity  94.4%;  Pred. No. 9.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           ||||||||| |||||||||
Db    3267 CGCGGCCGTCATCATCGC 3250


RESULT  10
MSU26405
LOCUS       MSU26405      778 bp    DNA             BCT       26-NOV-1996
DEFINITION  Mycobacterium smegmatis DNA polymerase I (polA) gene, partial cds.
ACCESSION   J26405
VERSION     J26405.1  GI:847764
KEYWORDS    .
SOURCE      Mycobacterium smegmatis.
  ORGANISM  Mycobacterium smegmatis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 246 to 308)
  AUTHORS   Gordhan,B.G., Andersen,S.J., De Meyer,A.R. and Mizrahi,V.
  TITLE     Construction by homologous recombination and phenotypic
            characterization of a DNA polymerase domain polA mutant of
            Mycobacterium smegmatis
  JOURNAL   Gene 178 (1-2), 125-130 (1996)
  MEDLINE   97080558
REFERENCE   2  (bases 1 to 778)
```

```
  AUTHORS   Mizrahi,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1995) Valerie J. Mizrahi, Molecular Biology Unit,
            South African Institute For Medical Research, Cnr. Hospital and De
            Korte Streets, Johannesburg, Gauteng 2000, South Africa
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Mycobacterium smegmatis"
                     /strain="mc2155"
                     /db_xref="taxon:1772"
                     /clone="pBG2"
     gene            1. .731
                     /gene="polA"
     CDS             <1. .731
                     /gene="polA"
                     /note="C-terminus of DNA polymerase domain"
                     /codon_start=3
                     /transl_table=11
                     /evidence=experimental
                     /product="DNA polymerase I"
                     /protein_id="AAC44593.1"
                     /db_xref="GI:847765"
                     /translation="IRDAFVVGEGYAELMTADYSQIEMRIMAHLSRDAGLIEAFNTGE
                     DLHSFVASRAFSVPIDEVTPELRRRVKAMSYGLAYGLSAYGWAQQLKISTEEAKVQME
                     QYFDRFGGVRDYLRDVVDQARKDGYTSTVLGRRRYLPELDSSNRQVREAAERAALNAP
                     IQGSAADIIKVAMINVDQAIKDAGLRSRILLQVHDELLFEVSEGERGELEQLVREHMG
                     NAYPLDVPLEVSVGYGRSWDAAAH"
     misc_feature    274. .279
                     /gene="polA"
                     /note="site of Km-r gene in polA755::aph mutant; Km-r
                     insertion in BglII site disrupts gene function; BglII
                     restriction site"
BASE COUNT      136 a    239 c    275 g    128 t
ORIGIN


  Query Match             88.9%;  Score 16;  DB 2;  Length 778;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATC 16
           ||||||||||||||||
Db     494 CGCGGCCGACATCATC 509


RESULT  11
A87607/c
LOCUS       A87607       1237 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9836089.
ACCESSION   A87607
VERSION     A87607.1  GI:6736247
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    386 g    221 t
ORIGIN


  Query Match             88.9%;  Score 16;  DB 5;  Length 1237;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CGGCCGACATCATCGC 18
           ||||||||||||||||
Db    1237 CGGCCGACATCATCGC 1222


RESULT  12
A89747/c
LOCUS       A89747       1237 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9832862.
ACCESSION   A89747
VERSION     A89747.1  GI:6738281
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    386 g    221 t
ORIGIN


  Query Match             88.9%;  Score 16;  DB 5;  Length 1237;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CGGCCGACATCATCGC 18
           ||||||||||||||||
Db    1237 CGGCCGACATCATCGC 1222


RESULT  13
MSGPOLA
LOCUS       MSGPOLA      3157 p    DNA             BCT       16-MAY-1996
DEFINITION  Mycobacterium tuberculosis DNA polymerase I (polA) gene, complete
            cds.
ACCESSION   L11920
VERSION     L11920.1  GI:416116
KEYWORDS    DNA polymerase I; polA gene.
SOURCE      Mycobacterium tuberculosis (strain H37Rv) DNA.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 2237 to 2843)
  AUTHORS   Mizrahi,V., Huberts,P., Dawes,S.S. and Dudding,L.R.
  TITLE     A PCR method for the sequence analysis of the gyrA, polA and rnhA
            gene segments from mycobacteria
  JOURNAL   Gene 136 (1-2), 287-290 (1993)
  MEDLINE   94124016
REFERENCE   2  (bases 1 to 3157)
  AUTHORS   Huberts,P. and Mizrahi,V.
  TITLE     Cloning and sequence analysis of the gene encoding the DNA
            polymerase I from Mycobacterium tuberculosis
  JOURNAL   Gene 164 (1), 133-136 (1995)
  MEDLINE   96060852
REFERENCE   3  (bases 1 to 3157)
  AUTHORS   Huberts,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1993) P. Huberts, University of the Witwatersrand
            Medical School, Haematology, 7 York Road, Parktown, Johannesburg
            2000, Transvaal, South Africa
COMMENT     On Nov 15, 1993 this sequence version replaced gi:309052.
FEATURES             Location/Qualifiers
     source          1. .3157
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
```

```
                     /clone="pPH1, pPH3"
                     /germline
     mat_peptide     302. .3013
                     /gene="polA"
                     /standard_name="DNA Polymerase I"
                     /note="Three possible positions for translational
                     initiation site: 302, 305, or 353"
                     /citation=[1]
                     /function="DNA polymerase"
                     /product="PolI"
     CDS             302. .3016
                     /gene="polA"
                     /standard_name="DNA polymerase I"
                     /function="DNA polymerase"
                     /note="Three possible positions for translational
                     initiation site"
                     /citation=[1]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="PolI"
                     /protein_id="AAB46393.1"
                     /db_xref="GI:416117"
                     /translation="MVTTASAPSEDRAKPTLMLLDGNSLAFRAFYALPAENFKTRGGL
                     TTNAVYGFTAMLINLLRDEAPTHIAAAFDVSRQTFRLQRYPEYKANRSSTPDEFAGQI
                     DITKEVLGALGITVLSEPGFEADDLIATLATQAENEGYRVLVVTGDRDALQLVSDDVT
                     VLYPRKGVSELTRFTPEAVVEKYGLTPRQYPDFAALRGDPSDNLPGIPGVGEKTAAKW
                     IAEYGSLRSLVDNVDAVRGKVGDALRANLASVVRNRELTDLVRDVPLAQTPDTLRLQP
                     WDRDHIHRLFDDLEFRVLRDRLFDTLAAAGGPEVDEGFDVRGGALAPGTVRQWLAEHA
                     GDGRRAGLTVVGTHLPHGGDATAMAVAAADGEGAYLDTATLTPDDDAALAAWLADPAK
                     PKALHEAKAAVHDLAGRGWTLEGVTSDTALAAYLVRPGQRSFTLDDLSLRYLRRELRA
                     ETPQQQQLSLLDDDDTDAETIQTTILRARAVIDLADALDAELARIDSTALLGEMELPV
                     QRVLAKMESAGIAVDLPMLTELQSQFGDQIRDAAEAAYGVIGKQINLGSPKQLQVVLF
                     DELGMPKTKRTKTGYTTDADALQSLFDKTGHPFLQHLLAHRDVTRLKVTVDGLLQAVA
                     ADGRIHTTFNQTIAATGRLSSTEPNLQNIPIRTDAGRRIRDAFVVGDGYAELMTADYS
                     QIEMRIMAHLSGDEGLIEAFNTGEDLHSFVASRAFGVPIDEVTGELRRRVKAMSYGLA
                     YGLSAYGLSQQLKISTEEANEQMDAYFARFGGVRDYLRAVVERARKDGYTSTVLGRRR
                     YLPELDSSNRQVREAAERAALNAPIQGSAADIIKVAMIQVDKALNEAQLASRMLLQVH
                     DELLFEIAPGERERVEALVRDKMGGAYPLDVPLEVSVGYGRSWDAAAH"
     gene            302. .3016
                     /gene="polA"
     stem_loop       3106. .3133
                     /citation=[2]
                     /function="Putative transcriptional stop"
BASE COUNT      552 a   1053 c   1054 g    498 t
ORIGIN


  Query Match           88.9%;  Score 16;  DB 2;  Length 3157;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATC 16
          ||||||||||||||||
Db   2779 CGCGGCCGACATCATC 2794


RESULT  14
AFANREA/c
LOCUS       AFANREA      4189 bp    DNA             BCT       03-AUG-1994
DEFINITION  Plasmid pTOM9 from Alcaligenes xylosoxidans nreA and nreB genes,
            complete cds.
ACCESSION   L31491
VERSION     L31491.1  GI:468276
KEYWORDS    nreA gene; nreB gene.
SOURCE      pTOM9 (plasmid pTOM9, kingdom Prokaryotae) DNA.
  ORGANISM  Plasmid pTOM9
            plasmids.
REFERENCE   1  (bases 1 to 4189)
  AUTHORS   Lemke,K.
  TITLE     Schwermetallresistenz in den zwei Alcaligenes-Staemmen A. eutrophus
            CH34 und A. xylosoxydans 31A: Transposonmutagenese, Klonierung und
            Sequenzierung. Cuvillier Verlag Goettingen; ISBN: 3-930340-28-3
  JOURNAL   Thesis (1994)
FEATURES             Location/Qualifiers
     source          1. .4189
                     /organism="Plasmid pTOM9"
                     /plasmid="pTOM9"
                     /specific_host="Alcaligenes xylosoxidans"
                     /db_xref="taxon:36546"
                     /note="strain 31A"
     CDS             176. .1063
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA72438.1"
                     /db_xref="GI:468277"
                     /translation="MCSNCYPPGSWLPICCHRSGCGTTDSSVRLDPDPVSIGDSAARY
                     RNTTHHSHSHMHIPVLPGRFDASRLRLLALVMVALLLLFTAMPSFAHGVAEGDKGYIQ
                     EIDGVHLLPFLYLGAKHMVTGYDHILFLLGVIFFLYRLKHIGLYVSLFALGHSTTMLL
                     GVYFNVGINSYVIDAIIGLSVVYKALDNMGAFQRWLGFQPNTKAATLIFGLFHGFGLS
                     TKILEYEISPNGLVPNLLAFNVGVEIGQLLALSAILIVMGFWRRTPSFLAHAYTANVA
                     MMSAGFILIGYQLTGLFVS"
     CDS             1074. .1844
                     /note="ORF2; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA72439.1"
                     /db_xref="GI:468278"
                     /translation="MYNTDIPARAELPTSARLIRSTAIAAAAATVILVTIVLPSEYAV
                     DPTGIGRVLGLTDMGKIKVQLAAEAKADAEKDAAAHAAQLAAANPVTAKPDSHGHAHG
                     PADTSKAANAHGHDHGGEAATGLAATPTAAPAAPAAQTVAAAAPSTADAAAGRSDKLN
                     FVLAPGQGLEVKMAMKAGARTSFDWTANGAVVNFDAHGDGNGRSVSYKKGKGAPSDTG
                     ELVAAFDGYHGWYWRNRTEKDVTMTLRTQGDYAEIKRM"
     -35_signal      2044. .2049
                     /note="putative"
     -10_signal      2067. .2072
                     /note="putative"
     gene            2103. .2372
                     /gene="nreA"
     CDS             2103. .2372
                     /gene="nreA"
                     /codon_start=1
                     /transl_table=11
                     /function="nickel resistance"
                     /evidence=experimental
                     /protein_id="AAA72440.1"
                     /db_xref="GI:468279"
                     /translation="MGVHTSHASIIKRLKRAEGHLRSIVVMMEEGRPCLAIAQQLQAV
                     ESAVTQAKKALVHDHIDHCLEEAVRDGTRPSDETLREFKSITKYL"
     gene            2376. .3800
                     /gene="nreB"
     CDS             2376. .3800
                     /gene="nreB"
                     /codon_start=1
                     /transl_table=11
                     /function="nickel resistance"
                     /evidence=experimental
                     /protein_id="AAA72441.1"
                     /db_xref="GI:468280"
                     /translation="MHARCPVVSGTPLACGLPSQFQAIGEISMLDVLKNRTYRHLFTA
                     QVIALIGTGLATVALGLLAYDLAGANAGSVLGTALAIKMVAYVGVAPVVGAYANRLPR
                     RAFLVAMDLIRAAVAISLPFVNEIWQIYILIFVLQSASAAFTPTFQATIPDVLRDEKD
                     YTRALSLSRLAYDMESLVSPMLAAALLSVISFHWLFGGTVVGFLISAALVVSVSLPQS
                     RAKEITGSIYAKTFRGISIYLRTPRLKGLLALNLAAAAASAMVIVNTVVYVQSKLQRP
                     STDVPFALAAFGCGSMLVALLLPRLLDKRPDRPAMLTGATIMAIGLAVGSLVAFAGGP
                     FEWHALLLTWFGIGVGYSMTQTPSGRLLKRSASSADRPAVFAAQFSLSHACWLITYPL
                     AGQLGARLGMGPTAAVMSVITVVGLFMAWKLWPSRDEDVIAHEHPDEEGMHVHLAEGE
                     RTGERSHRHAFTIDDQHPRWPDAH"
     terminator      3829. .3856
                     /note="putative"
BASE COUNT      770 a   1318 c   1236 g    865 t
ORIGIN


  Query Match           88.9%;  Score 16;  DB 1;  Length 4189;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CGGCCGACATCATCGC 18
           ||||||||||||||||
Db    1016 CGGCCGACATCATCGC 1001


RESULT  15
MTCY01B2
LOCUS       MTCY01B2    35938 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 72/162.
ACCESSION   Z95554 AL123456
VERSION     Z95554.1  GI:3261771
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 35938)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 35938)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2113893.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .35938
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .1108
                     /note="fragment designated v046. Does not represent a
                     physical clone"
     RBS             88. .93
                     /note="possible RBS, ggaaga, for Rv1609"
     gene            101. .1651
                     /gene="trpE"
     CDS             101. .1651
                     /gene="trpE"
                     /note="Rv1609, (MTCY01B2.01-MTV046.07), trpE, len: 516.
                     Function: anthranilate synthase component I (EC 4.1.3.27).
                     FASTA best: TRPE_CLOTM P14953 anthranilate synthase
                     componenti (ec (494 aa), E(): 0, 42.6% identity in 498 aa
                     overlap.Some similarity to MTCY253.35, 6.3e-17;
                     MTCY07D11.11c, 5.7e-15"
                     /codon_start=1
                     /transl_table=11
                     /product="trpE"
                     /protein_id="CAB08903.1"
                     /db_xref="GI:3261772"
                     /db_xref="SPTREMBL:O06127"
                     /translation="MHADLAATTSREDFRLLAAEHRVVPVTRKVLADSETPLSAYRKL
                     AANRPGTFLLESAENGRSWSRWSFIGAGAPTALTVREGQAVWLGAVPKDAPTGGDPLR
                     ALQVTLELLATADRQSEPGLPPLSGGMVGFFAYDMVRRLERLPERAVDDLCLPDMLLL
                     LATDVAAVDHHEGTITLIANAVNWNGTDERVDWAYDDAVARLDVMTAALGQPLPSTVA
                     TFSRPEPRHRAQRTVEEYGAIVEYLVDQIAAGEAFQVVPSQRFEMDTDVDPIDVYRIL
                     RVTNPSPYMYLLQVPNSDGAVDFSIVGSSPEALVTVHEGWATTHPIAGTRWRGRTDDE
                     DVLLEKELLADDKERAEHLMLVDLGRNDLGRVCTPGTVRVEDYSHIERYSHVMHLVST
                     VTGKLGEGRTALDAVTACFPAGTLSGAPKVRAMELIEEVEKTRRGLYGGVVGYLDFAG
                     NADFAIAIRTALMRNGTAYVQAGGGVVADSNGSYEYNEARNKARAVLNAIAAAETLAA
                     PGANRSGC"
     source          1008. .35698
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y01B2"
     gene            1641. .2348
                     /gene="Rv1610"
     CDS             1641. .2348
                     /gene="Rv1610"
                     /note="Rv1610, (MTCY01B2.02), len: 235. Possible
                     membraneprotein"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1610"
                     /protein_id="CAB08904.1"
                     /db_xref="GI:2113894"
                     /db_xref="SPTREMBL:O06128"
                     /translation="MAANAGSVRPNRRARPMIGIAQLLLVVAAGALWMAARLPWVVIG
                     SFDELGPPKEVTLTGASWSTALLPLALLMLAAAVAALAVRGWPLRALAVLLAAASFAV
                     GYLGISLWVVPDVAARGADLAHVPVVTLVGSARHYWGAVAAVLAAVCALLAAVFLMSS
                     AAIRGSAGEDMARYAAPRARRSIARRQHSNAAGRAAPQDDGPDMGPRMSERMIWEALD
                     EGRDPTDREQESDTEGR".
     RBS             2423. .2426
                     /note="possible RBS, AAGG, for Rv1611"
     gene            2438. .3256
                     /gene="trpC"
     CDS             2438. .3256
                     /gene="trpC"
                     /note="Rv1611, (MTCY01B2.03), len: 272. Function:
                     probableindole-3-glycerol phosphate synthase (EC
                     4.1.1.48). Contains indole-3-glycerol phosphate synthase
                     signature (PS00614).FASTA best: SLR0546 SYNECHOCYSTIS SP
                     Q55508 HYPOTHETICAL33.0 KD PROTEIN. (295 aa) opt:  26;
                     E(): 7.6e-32; (44.2% identity in 265 aa overlap) aslo
                     similar to eg TRPC_AZOBR P26938(262 aa) opt: 596 z-score:
                     665.0 E(): 4.8e-30 (43.8% identity in 258 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="trpC"
                     /protein_id="CAB08905.1"
                     /db_xref="GI:2113895"
                     /db_xref="SPTREMBL:O06129"
                     /translation="MSPATVLDSILEGVRADVAAREASVSLSEIKAAAAAAPPPLDVM
                     AALREPGIGVIAEVKRASPSAGALATIADPAKLAQAYQDGGARIVSVVTEQRRFQGSL
                     DDLDAVRASVSIPVLRKDFVVQPYQIHEARAHGADMLLLIVAALEQSVLVSMLDRTES
                     LGMTALVEVHTEQEADRALKAGAKVIGVNARDLMTLDVDRDCFARIAPGLPSSVIRIA
                     ESGVRGTADLLAYAGAGADAVLVGEGLVTSGDPRAAVADLVTAGTHPSCPKPAR"
     misc_feature    2597. .2641
                     /gene="trpC"
```

```
                     /note="PS00614 Indole-3-glycerol phosphate synthase
                     signature"
     gene            3325. .4557
                     /gene="trpB"
     CDS             3325. .4557
                     /gene="trpB"
                     /note="Rv1612, (MTCY01B2.04), trpB, len: 410. Function:
                     probable tryptophan synthase beta chain (EC 4.2.1.20).
                     FASTAbest: TRPB_CAUCR P12290 tryptophan synthase beta
                     chain (406 aa) opt: 1662; E(): 0; (60.6% identity in 404
                     aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="trpB"
                     /protein_id="CAB08906.1"
                     /db_xref="GI:2113896"
                     /db_xref="SWISS-PROT:O08376"
                     /translation="MSAAIAEPTSHDPDSGGHFGGPSGWGGRYVPEALMAVIEEVTAA
                     YQKERVSQDFLDDLDRLQANYAGRPSPLYEATRLSQRAGSARIFLKREDLNHTGSHKI
                     NNVLGQALLARRMGKTRVIAETGAGQHGVATATACALLGLDCVIYMGGIDTARQALNV
                     ARMRLLGAEVVAVQTGSKTLKDAINEAFRDWVANADNTYYCFGTAAGPHPFPTMVRDF
                     QRIIGMEARVQIQGQAGRLPDAVVACVGGGSNAIGIFHAFLDDPGVRLVGFEAAGDGV
                     ETGRHAATFTAGSPGAFHGSFSYLLQDEDGQTIESHSISAGLDYPGVGPEHAWLKEAG
                     RVDYRPITDSEAMDAFGLLCRMEGIIPAIESAHAVAGALKLGVELGRGAVIVVNLSGR
                     GDKDVETAAKWFGLLGND"
     misc_feature    3604. .3633
                     /gene="trpB"
                     /note="PS00168 Tryptophan synthase beta ch in
                     pyridoxal-phosphate attachment site"
     gene            4557. .5369
                     /gene="trpA"
     CDS             4557. .5369
                     /gene="trpA"
                     /note="Rv1613, (MTCY01B2.05), trpA, len: 270. Function:
                     probable tryptophan synthase alpha chain (EC 4.2.1.20).
                     FASTAbest: TRPA_THETH P16608 t yptophan synthase alpha
                     chain (271 aa) opt:  601; E(): 2.2e-30; (43.0% identity in
                     251 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="trpA"
                     /protein_id="CAB08890.1"
                     /db_xref="GI:2113897"
                     /db_xref="SPTREMBL:O06130"
                     /translation="MVAVEQSEASRLGPVFDSCRANNRAALIGYLPTGYPDVPASVAA
                     MTALVESGCDIIEVGVPYSDPGMDGPTIARATEAALRGGVRVRDTLAAVEAISIAGGR
                     AVVMTYWNPVLRYGVDAFARDLAAAGGLGLITPDLIPDEAQQWLAASEEHRLDRIFLV
                     APSSTPERLAATVEASRGFVYAASTMGVTGARDAVSQAAPELVGRVKAVSDIPVGVGL
                     GVRSRAQAAQIAQYADGVIVGSALVTALTEGLPRLRALTGELAAGVRLGMSA"
     gene            5369. .6775
                     /gene="lgt"
     CDS             5369. .6775
                     /gene="lgt"
                     /note="Rv1614, (MTCY01B2.06), len: 468. Unknown,
                     possiblelgt/umpA, similar to many prolipoprotein
                     diacylglyceryl transferase (EC 2.4.99.-). FASTA best:
                     LGT_STAAU P52282 prolipoprotein diacylglyceryl transfer
                     (279 aa) opt:  289; E():3.6e-09; (31.5% identity in 257 aa
                     overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="lgt"
                     /protein_id="CAB08891.1"
                     /db_xref="GI:2113898"

  Query Match             88.9%;  Score 16;  DB 1;  Length 35938;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATC 16
           ||||||||||||||||
Db   25340 CGCGGCCGACATCATC 25355
```

Search completed: June 22, 2000, 17:58:40
Job time: 10789 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:12:46 ; Search time 458.59 Seconds
(without alignments)
9.820 Million cell updates/sec

Title: US-09-362-485-15
Perfect score: 18
Sequence: 1 GCCCCGATGAGCGAAGTC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	18	100.0	18	1	V49516	Mycobacterium sp.
2	18	100.0	18	1	V49614	AlaDH derived olig
3	18	100.0	1245	1	V49626	Mycobacterium tube
4	18	100.0	1260	1	V49510	Mycobacterium sp.
5	18	100.0	1260	1	V49625	Mycobacterium tube
c 6	16	88.9	4328	1	V74590	Staphylococcus aur
7	14.8	82.2	682	1	V49511	Mycobacterium mari
c 8	14.8	82.2	7011	1	V20464	Human L-myc oncoge
c 9	14.4	80.0	375	1	V20165	Probe (66) for mic
10	14.4	80.0	4348	1	V29008	Parathyroid hormon
c 11	14	77.8	2014	1	Q36619	CMV Colburn cDNA.
c 12	13.8	76.7	104	1	V88832	EST clone HR593. N
13	13.8	76.7	1125	1	N91423	Sequence of heat-r
14	13.8	76.7	1251	1	V81736	Rhodococcus sp. st
15	13.8	76.7	5533	1	Q55131	Sequence encoding
16	13.8	76.7	5535	1	Q55132	Sequence encoding
17	13.8	76.7	5535	1	T58971	Rhodococcus rhodoc
18	13.8	76.7	5535	1	X15615	DNA encoding enzym
19	13.8	76.7	5535	1	X15616	DNA encoding an en
c 20	13.8	76.7	40352	1	V02032	MAGE-B cluster DNA
c 21	13.4	74.4	489	1	V24789	H. pylori ORF 09ap
c 22	13.4	74.4	528	1	V24938	H. pylori secreted
c 23	13.4	74.4	762	1	V25110	H. pylori secreted
c 24	13.4	74.4	982	1	V52040	Helicobacter polyp
25	13.4	74.4	1037	1	Q04705	USP-Promoter-casse
26	13.4	74.4	1085	1	Q04703	Legumin-signalpept
27	13.4	74.4	1110	1	X51701	DNA encoding a hum
28	13.4	74.4	1160	1	Q04706	USP-signalpeptide
29	13.4	74.4	1166	1	Q04704	USP-Promoter-casse
30	13.4	74.4	1303	1	T39337	Plasmid pTS88 (Eco
31	13.4	74.4	3201	1	Q14529	pPS029 Bt ICP codi
32	13.4	74.4	3201	1	Q15144	pVE36 Bt ICP codin
c 33	13.4	74.4	4946	1	T59531	T-DNA of plasmid p
c 34	13.4	74.4	5349	1	V23239	T-DNA of pTTS24. D
35	13.4	74.4	5864	1	T39339	Plasmid pTC0113 T-
c 36	13.4	74.4	5864	1	T39339	Plasmid pTC0113 T-
37	13.4	74.4	6548	1	T39336	Plasmid pTS174 use
38	13.4	74.4	6548	1	T61394	Plasmid pTS172. Pr
39	13.4	74.4	7566	1	Q42160	Plasmid pPS0212 co
40	13.4	74.4	7639	1	Q42159	Plasmid pJD884 con
c 41	13.2	73.3	516	1	Q28828	Recombinant DNA en
c 42	13.2	73.3	572	1	Q43458	Sequence of xylana
43	13.2	73.3	681	1	Q25013	Rice allergen gene
44	13.2	73.3	681	1	Q47086	O. sativa allergen
c 45	13.2	73.3	1123	1	Q74098	Humicola xylanase

ALIGNMENTS

RESULT 1
V49516
ID V49516 standard; DNA; 18 BP.
AC V49516;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-F5.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 4 A; 6 C; 5 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||||||||||||||||
Db 1 GCCCCGATGAGCGAAGTC 18

RESULT 2
V49614
ID V49614 standard; DNA; 18 BP.
AC V49614;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-F5.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 4 A; 6 C; 6 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||||||||||||||||
Db 1 GCCCCGATGAGCGAAGTC 18

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||||||||||||||||
Db 457 GCCCCGATGAGCGAAGTC 474

RESULT 4
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 413 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||||||||||||||||
Db 472 GCCCCGATGAGCGAAGTC 489

RESULT 5
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN 1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||||||||||||||||
Db 472 GCCCCGATGAGCGAAGTC 489

RESULT 6
V74590/c
ID V74590 standard; DNA; 4328 BP.
AC V74590;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #279.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 1081. .1140
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 2881. .2940
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1101-1103; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 4328 BP; 1282 A; 776 C; 626 G; 1523 T;

Query Match 88.9%; Score 16; DB 1; Length 4328;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCCCGATGAGCGAAGT 17
||||||||||||||||
Db 2057 CCCCGATGAGCGAAGT 2042

RESULT 7
V49511
ID V49511 standard; DNA; 682 BP.
AC V49511;
DT 20-OCT-1998 (first entry)
DE Mycobacterium marinum Mar3 DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium murinum.
FH Key Location/Qualifiers
FT CDS 1. .682
FT /*tag= a
FT /codon_start= 3
FT /product= "Mar3"
FT /note= "Alanine dehydrogenase"
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
DR P-PSDB; W64481.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Claim 1; Page 34-35; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3
CC isolated from Mycobacterium murinum. This protein is used to diagnose
CC tuberculosis and other mycobacterial infections (including 'swimmers'
CC disease', caused by M. marinum, a fish pathogen) in humans or animals.
CC The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 82.2%; Score 14.8; DB 1; Length 682;
Best Local Similarity 88.9%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||||| |||||||| |||
Db 354 GCCCCCATGAGCGAGGTC 371

RESULT 8
V20464/c
ID V20464 standard; DNA; 7011 BP.
AC V20464;
DT 17-JUN-1998 (first entry)
DE Human L-myc oncogene.
KW Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW cancer; antisense oligonucleotide; L-myc; ds.
OS Homo sapiens.
PN US5734039-A.
PD 31-MAR-1998.

PF 15-SEP-1994; 306691.
PR 15-SEP-1994; US-306691.
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Calabretta B, Skorski T;
DR WPI; 98-229882/20.
PT Anticancer composition comprising two anti-sense oligo:nucleotide(s)
PT - targetting cytoplasmic and nuclear oncogene(s)
PS Claim 1; Column 119-126; 92pp; English.
CC The present sequence represents an oncogene from the present invention.
CC The present invention describes a composition which comprises two
CC antisense oligonucleotides. The first oligonucleotide is specific for a
CC cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC c-rel, c-vav, c-ski, c-spi, cyclin D1, PML/RAR alpha , AML1/MTG8,
CC E2A/prl and ALL-1/AF-4. The composition is used for treating cancer.
CC The combination of antisense oligonucleotides has synergistically
CC enhanced ability to inhibit growth of cancer cells.
SQ Sequence 7011 BP; 1561 A; 1825 C; 1924 G; 1701 T;

Query Match 82.2%; Score 14.8; DB 1; Length 7011;
Best Local Similarity 88.9%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
|||| ||| |||||||||
Db 2735 GCCCAGATCAGCGAAGTC 2718

RESULT 9
V20165/c
ID V20165 standard; DNA; 375 BP.
AC V20165;
DT 26-JUN-1998 (first entry)
DE Probe (66) for microbial genes induced during host infection.
KW Probe; identification; microbial gene; pathogenic microorganism;
KW host infection; virulence gene; vaccine; antimicrobial agent; ss.
OS Salmonella typhimurium.
PN WO9744487-A1.
PD 27-NOV-1997.
PF 16-MAY-1997; U08208.
PR 17-MAY-1996; US-651155.
PA (REGC) UNIV CALIFORNIA.
PI Conner CP, Heithoff DM, Mahan MJ;
DR WPI; 98-018538/02.
PT Identification of microbial coding sequences - for use in vaccines
PT against virulent pathogenic microorganisms
PS Claim 3; Page 20; 172pp; English.
CC The present sequence, which is derived from a Salmonella
CC typhimurium gene specifically induced during host infection, can be
CC used as a probe for the identification of microbial genes
CC specifically induced in a pathogenic microorganism during host
CC infection, i.e. virulence genes. The products of the identified
CC virulence genes provide targets for the development of vaccines or
CC antimicrobial agents. The vaccines can be used to protect a
CC mammalian host against the pathogenesis of virulent microorganisms.
SQ Sequence 375 BP; 95 A; 83 C; 99 G; 98 T;

Query Match 80.0%; Score 14.4; DB 1; Length 375;
Best Local Similarity 93.8%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCCGATGAGCGAAGT 17
||||| ||||||||||
Db 82 CCCCGGTGAGCGAAGT 67

RESULT 10
V29008
ID V29008 standard; DNA; 4348 BP.
AC V29008;
DT 28-AUG-1998 (first entry)
DE Parathyroid hormone-related protein (PTH-rP) gene promoter.
KW Parathyroid hormone-related protein; PTH-rP; promoter; human;
KW tumour; metastasis; breast cancer; lung cancer; therapy;
KW bone formation; ss.
OS Homo sapiens.
PN WO9807844-A2.
PD 26-FEB-1998.
PF 22-AUG-1997; U14836.
PR 21-AUG-1997; US-915868.
PR 23-AUG-1996; US-025215.
PA (OSTE-) OSTEOSCREEN.
PI Gallwitz WE, Mundy GR;
DR WPI; 98-169153/15.
PT Screening compounds for alteration of parathyroid hormone-related
PT protein production - potentially useful for treating cancer and for
PT stimulating bone formation and repair
PS Disclosure; Page 33-36; 51pp; English.
CC This polynucleotide comprises the promoter region of the
CC parathyroid hormone-related protein (PTH-rP) gene. 3 Promoters are
CC contained within the 4.3 kb genomic DNA upstream from the start
CC site of the PTH-rP gene, located on chromosome 12. Promoter P1 is
CC 5' of exon 1A (the upstream TATA element); P2 is 5' of exon 1C (the
CC midregion GC rich element); and P3 is 5' of exon 2 (downstream
CC TATA element). 3 Isoforms of PTH-rP arise from alternative 3'
CC splicing; most tumours expressing PTH-rP express PTH-rP 1-141
CC driven by P3. The invention provides a cell-based assay for
CC identifying compounds that affect PTH-rP production in mammalian
CC cells. The method involves: (i) transfecting the cells with an
CC expression construct containing the PTH-rP promoter and a reporter
CC gene under control of the promoter; (ii) contacting the cells with
CC the test compound; and (iii) identifying compounds that alter
CC reporter gene expression. The mammalian cells may be human tumour
CC cells, such as breast cancer cells (e.g. MDA-MB-231 cells), lung
CC cancer cells (e.g. RWGT2 cells) and bone cells (e.g. MC3T3, MG-63,
CC U2OS, UMR-106, ROS17/2.8 and SAOS-2). Identified compounds are
CC used: (a) for inhibiting tumour growth (PTH-rP is involved in
CC metastasis to the bone, so inhibiting it will halt progressive bone
CC loss and hypercalcaemia associated with cancers); and (b) for
CC stimulating bone formation and repair (PTH-rP is essential for
CC normal cartilage differentiation and endo-chondral bone formation,
CC and has an anabolic effect on bone).
SQ Sequence 4348 BP; 1059 A; 996 C; 1098 G; 1195 T;

Query Match 80.0%; Score 14.4; DB 1; Length 4348;
Best Local Similarity 93.8%; Pred. No. 34;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAG 16
|||||||||||||| |
Db 2219 GCCCCGATGAGCGAGG 2234

RESULT 11
Q36619/c
ID Q36619 standard; cDNA; 2014 BP.
AC Q36619;
DT 02-JUN-1993 (first entry)
DE CMV Colburn cDNA.
KW Cytomegalovirus; CMV; Colburn; assembly protein; gene; herpes; virus;
KW proteinase; nested 3' co-terminal genes; APNG1; APNG.5; precursor;
KW proteolytic cleavage; conserved; motif; maturation; cleavage site;
KW autoprocessing; enzyme release; ss.
OS Simian cytomegalovirus.
FH Key Location/Qualifiers
FT cds 175. .2004
FT /*tag= a
FT signal_peptide 175. .231
FT /*tag= b
FT mat_peptide 232. .2001
FT /*tag= c

```
FT   mrna            232. .726
FT                   /*tag= d
FT                   /note= "ORF APNG1"
FT   mrna            727. .1071
FT                   /*tag= e
FT                   /note= "ORF APNG.7"
FT   mrna            1072. .1275
FT                   /*tag= f
FT                   /note= "ORF APNG.5"
FT   mrna            1276. .2004
FT                   /*tag= g
FT                   /note= "ORF APNG.4"
PN   WO9301291-A.
PD   21-JAN-1993.
PF   02-JUL-1992; U05513.
PR   05-JUL-1991; US-725308.
PR   27-NOV-1991; US-798776.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Gibson DW,  Welch AR;
DR   WPI; 93-045494/05.
DR   P-PSDB; R32131.
PT   New herpes virus proteinase cleaving virus assembly protein
PT   precursor - for measuring the activity of a herpes virus
PT   proteinase and for diagnosis or treatment of a herpes virus
PT   infection, free of infectious virion DNA
PS   Disclosure; Fig 1; 58pp; English.
CC   The sequence given shows the nucleotide sequence of the
CC   cytomegalovirus (CMV)-Colburn genomic region which contains the
CC   assembly protein gene at the end of a 1770 bp open reading frame
CC   (ORF).  The assembly protein of herpes viruses is maturationally
CC   processed by a herpes virus-encoded proteinase.  The proteinase is
CC   a member of a family of four nested 3' co-terminal genes which
CC   includes the assembly protein itself.  Each of the genes appears to
CC   be transcribed into separate mRNAs. The ORF designated APNG1 denotes
CC   the beginning of the coding sequence of the proteinase gene, and the
CC   ORF designated APNG.5 denotes the beginning of the coding sequence of
CC   the precursor assembly protein gene. It appears that proteolytic
CC   cleavage occurs at a number of locations in this family of gene
CC   products.  The assembly protein precusor is cleaved between the Ala557
CC   and Ser558 of the CMV Colburn gene product (amino acids being number
CC   from the first amino acid of the mature protein).  The cleavage site
CC   in herpes virus assembly protein precursors have the conserved motif
CC   of aa1-aa2-Ala-aa3, where aa1 is Ala or Leu, aa2 is a polar amino acid
CC   and aa3 is Ser, Asn or Val.  This motif is referred to as the assembly
CC   protein maturation cleavage site.  Another putative cleavage site
CC   occurs after the Ala residue of the sequence Tyr-Val-Lys-Ala, which
CC   occurs at amino acids 246 to 249 in the CMV Colburn APNG1 gene product.
CC   This site has been found to be used in vivo for autoprocessing of the
CC   proteinase molecule.  This cleavage site is referred to as the enzyme
CC   release cleavage site.  Both the maturation and enzyme release cleavage
CC   sites are highly conserved among herpes viruses.
SQ   Sequence   2014 BP;    371 A;    594 C;    655 G;    394 T;
```

```
  Query Match            77.8%;  Score 14;  DB 1;  Length 2014;
  Best Local Similarity  100.0%;  Pred. No. 54;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGA 14
          ||||||||||||||
Db   1199 GCCCCGATGAGCGA 1186
```

```
RESULT  12
V88832/c
ID   V88832 standard; cDNA; 104 BP.
AC   V88832;
DT   12-FEB-1999  (first entry)
DE   EST clone HR593.
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS   Homo sapiens.
PN   WO9845437-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06956.
PR   10-APR-1997; US-837312.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070078/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
PS   Claim 1; Page 527; 641pp; English.
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
SQ   Sequence   104 BP;    36 A;    30 C;    18 G;    20 T;
```

```
  Query Match            76.7%;  Score 13.8;  DB 1;  Length 104;
  Best Local Similarity  88.2%;  Pred. No. 57;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          |||| |||||| |||||
Db     50 GCCCAGATGAGAGAAGT 34
```

```
RESULT  13
N91423
ID   N91423 standard; DNA; 1125 BP.
AC   N91423;
DT   01-FEB-1991  (first entry)
DE   Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt.
DE   2 MD or less
KW   Enzyme; ds.
OS   Bacillus stearothermophilus IFO 12550.
PN   J01043194-A.
PD   15-FEB-1989.
PF   10-AUG-1987; 200524.
PR   10-AUG-1987; JP-200524.
PA   (NIRA) Unitika Kk.
DR   WPI; 89-096096/13.
PT   Recombinant plasmid for transforming Escherichia coli -
PT   obtd. by connecting heat resistant alanine dehydrogenase gene to
PT   vector plasmid
PS   Disclosure; Fig 2, p693 ; 9pp; Japanese.
CC   A recombinant plasmid contg. heat-resistant alanine dehydrogenase (AH)
CC   gene with mol. wt. 2 or less MD is claimed.  Cells transformed with the
CC   vector produce high levels of heat-resistant AH.  Transformed E. coli
CC   is useful as a clinical inspection reagent.
SQ   Sequence   1125 BP;    246 A;    295 C;    367 G;    217 T;
```

```
  Query Match            76.7%;  Score 13.8;  DB 1;  Length 1125;
  Best Local Similarity  88.2%;  Pred. No. 66;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 CCCCGATGAGCGAAGTC 18
          | |||||||| ||||||
Db    395 CGCCGATGAGTGAAGTC 411
```

```
RESULT  14
V81736
ID   V81736 standard; DNA; 1251 BP.
AC   V81736;
DT   25-FEB-1999  (first entry)
DE   Rhodococcus sp. strain IGTS8 desulphurisation activity ORF-3(R) DNA.
KW   Rhodococcus sp. strain IGTS8; Sphingomonas sp. strain AD109;
KW   desulfurization activity; desulphurisation; DSZ; pseudomonad;
KW   sulphur-containing heterocycle; fossil fuel; petroleum; oxidation; ss.
OS   Rhodococcus sp.
FH   Key             Location/Qualifiers
FT   CDS             1. .1251
FT                   /*tag= a
FT                   /note= "no stop codon given"
PN   WO9845447-A1.
PD   15-OCT-1998.
PF   03-APR-1998; U06691.
PR   05-MAY-1997; US-851088.
PR   07-APR-1997; US-835185.
PA   (ENER-) ENERGY BIOSYSTEMS CORP.
PI   Childs JD, Darzins A, Monticello DJ, Squires CH,
PI   Xi L;
DR   WPI; 99-009331/01.
DR   P-PSDB; W89243.
PT   New recombinant pseudomonad containing heterologous desulphurisation
PT   enzyme genes - used for removing sulphur-containing heterocycles
PT   from fossil fuels, especially petroleum, and for oxidation of
PT   organic substrates
PS   Claim 5; Fig 3; 130pp; English.
CC   The present sequence encodes a Rhodococcus sp. strain IGTS8
CC   desulphurisation activity ORF. The present invention describes a
CC   recombinant pseudomonad comprising a heterologous nucleic acid encoding
CC   a desulphurisation enzyme. The recombinant pseudomonads, or enzyme
CC   preparations derived from them, are used: (i) to desulphurise
CC   carbonaceous materials containing organosulphur compounds, preferably
CC   fossil fuels and especially petroleum or its distillate fractions, but
CC   also coal; and (ii) more generally to oxidise organic compounds,
CC   particularly organosulphur compounds or optionally substituted indoles.
CC   The recombinant pseudomonad oxidises carbon-sulphur bonds in
CC   heterocycles that are refractory to hydrodesulphurisation treatments,
CC   particularly dibenzothiophene (DBT) and its derivatives. The recombinant
CC   pseudomonads can be produced at relatively low cost and grow well in the
CC   desulphurisation system, allowing them to compete effectively with any
CC   biological contaminants. High concentrations of the recombinant
CC   pseudomonads can be tolerated without complicating recovery of
CC   desulphurised products.
SQ   Sequence   1251 BP;    252 A;    435 C;    373 G;    191 T;

  Query Match            76.7%;  Score 13.8;  DB 1;  Length 1251;
  Best Local Similarity  88.2%;  Pred. No. 67;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          |||||||||| |||| |
Db    301 GCCCCGATGATCGAACT 317


RESULT  15
Q55131
ID   Q55131 standard; DNA; 5533 BP.
AC   Q55131;
DT   02-AUG-1994  (first entry)
DE   Sequence encoding enzymes which desulphurise a fossil fuel.
KW   Enzyme; biocatalyst; fossil fuel; oxidation; cleavage;
KW   organosulphur compounds; coal; ds.
OS   Rhodococcus rhodochrous.
FH   Key             Location/Qualifiers
FT   cds             790. .2151
FT                   /*tag= a
FT                   /label= ORF 1
FT   cds             3256. .4506
FT                   /*tag= b
FT                   /label= ORF 3
PN   WO9401563-A.
PD   20-JAN-1994.
PF   09-JUL-1993; U06497.
PR   10-JUL-1992; US-911845.
PA   (ENER-) ENERGY BIOSYSTEMS CORP.
PI   Denome SA,  Kovacevich BR,  Piddington CS,  Rambosek J;
PI   Young KD;
DR   WPI; 94-035068/04.
DR   P-PSDB; R47872, R47873.
PT   DNA encoding a bio catalyst which desulphurises fossil fuels -
PT   obtd. from Rhodococcus rhodochrous bacteria, used to produce
PT   microorganisms which degrade organic sulphur cpds.
PS   Claim 39; Page 67-72; 104pp; English.
CC   Microorganisms transformed with the DNA can be used to produce
CC   the enzymes/biocatalysts for the selective oxidative cleavage of
CC   carbon-sulphur bonds for desulphurisation of fossil fuels which
CC   contain organosulphur compounds.
SQ   Sequence   5533 BP;    1027 A;    1808 C;    1729 G;    969 T;

  Query Match            76.7%;  Score 13.8;  DB 1;  Length 5533;
  Best Local Similarity  88.2%;  Pred. No. 74;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          |||||||||| |||| |
Db   3556 GCCCCGATGATCGAACT 3572


Search completed: June 22, 2000, 15:12:48
Job time: 18094 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:21:54 ; Search time 341.15 Seconds
(without alignments)
6.858 Million cell updates/sec

Title: US-09-362-485-15
Perfect score: 18
Sequence: 1 GCCCCGATGAGCGAAGTC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14.8	82.2	7011	2	US-08-306-691B-42	Sequence 42, Appl
	2	14.4	80.0	4348	4	US-08-915-868-1	Sequence 1, Appli
c	3	14	77.8	2014	1	US-07-798-776-1	Sequence 1, Appli
c	4	14	77.8	2014	5	US-08-251-288A-1	Sequence 1, Appli
	5	13.8	76.7	1251	4	US-08-851-088-5	Sequence 5, Appli
	6	13.8	76.7	5535	1	US-08-089-755A-1	Sequence 1, Appli
	7	13.8	76.7	5535	1	US-08-089-755A-4	Sequence 4, Appli
	8	13.8	76.7	5535	1	US-08-421-754-1	Sequence 1, Appli
	9	13.8	76.7	5535	1	US-08-421-754-4	Sequence 4, Appli
	10	13.8	76.7	5535	3	US-08-421-791-1	Sequence 1, Appli
	11	13.8	76.7	5535	3	US-08-421-791-4	Sequence 4, Appli
c	12	13.8	76.7	40352	5	US-08-846-111D-15	Sequence 15, Appl
	13	13.4	74.4	1303	5	US-08-894-440-2	Sequence 2, Appli
	14	13.4	74.4	3200	1	US-08-453-104-23	Sequence 23, Appl
	15	13.4	74.4	3200	3	US-08-694-824-23	Sequence 23, Appl
	16	13.4	74.4	3201	1	US-08-453-104-22	Sequence 22, Appl
	17	13.4	74.4	3201	3	US-08-694-824-22	Sequence 22, Appl
	18	13.4	74.4	5864	5	US-08-894-440-4	Sequence 4, Appli
c	19	13.4	74.4	5864	5	US-08-894-440-4	Sequence 4, Appli
	20	13.4	74.4	6548	5	US-08-894-440-1	Sequence 1, Appli
	21	13.4	74.4	7566	4	US-08-232-016-23	Sequence 23, Appl
	22	13.4	74.4	7639	4	US-08-232-016-22	Sequence 22, Appl
c	23	13.4	74.4	7811	4	US-08-549-680A-5	Sequence 5, Appli
c	24	13.2	73.3	489	1	US-08-119-169A-7	Sequence 7, Appli
c	25	13.2	73.3	1123	1	US-08-458-023B-3	Sequence 3, Appli
	26	13.2	73.3	1737	3	US-08-750-703-2	Sequence 2, Appli
	27	13.2	73.3	1740	3	US-08-750-703-1	Sequence 1, Appli
	28	13.2	73.3	2639	1	US-07-952-817-8	Sequence 8, Appli
	29	13.2	73.3	2639	7	5210025-1	Patent No. 5210025
c	30	13.2	73.3	3090	2	US-08-102-942A-3	Sequence 3, Appli
	31	13.2	73.3	4032	1	US-08-126-587C-8	Sequence 8, Appli
	32	13.2	73.3	5859	1	US-08-312-387B-1	Sequence 1, Appli
	33	13.2	73.3	5859	1	US-08-312-387B-7	Sequence 7, Appli
	34	13.2	73.3	5859	2	US-08-683-426-1	Sequence 1, Appli
	35	13.2	73.3	5859	2	US-08-683-426-7	Sequence 7, Appli
	36	13.2	73.3	5859	2	US-08-683-458-1	Sequence 1, Appli
	37	13.2	73.3	5859	2	US-08-683-458-7	Sequence 7, Appli
	38	13.2	73.3	5859	4	US-08-878-360-1	Sequence 1, Appli
	39	13.2	73.3	5859	4	US-08-878-360-7	Sequence 7, Appli
	40	13.2	73.3	7216	3	US-08-750-703-3	Sequence 3, Appli
	41	13.2	73.3	8779	3	US-08-750-703-4	Sequence 4, Appli
c	42	13.2	73.3	15378	5	US-08-785-420-1	Sequence 1, Appli
c	43	13.2	73.3	43795	5	US-08-742-185-101	Sequence 101, App
	44	13	72.2	810	1	US-08-008-216-2	Sequence 2, Appli
	45	13	72.2	810	1	US-08-459-569-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-306-691B-42/c
; Sequence 42, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-306-691B-42

Query Match 82.2%; Score 14.8; DB 2; Length 7011;
Best Local Similarity 88.9%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 GCCCCGATGAGCGAAGTC 18
          |||| ||| |||||||||
Db   2735 GCCCAGATCAGCGAAGTC 2718


RESULT   2
US-08-915-868-1
; Sequence 1, Application US/08915868
; Patent No. 5914233
;  GENERAL INFORMATION:
;    APPLICANT:  Mundy, Gregory R.
;    APPLICANT:  Gallwitz, Wolf
;    TITLE OF INVENTION:  SCREENING ASSAY FOR THE IDENTIFICATION
;    TITLE OF INVENTION:  OF AGENTS WHICH INHIBIT CANCER
;    TITLE OF INVENTION:  METASTASIS TO BONE
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/915,868
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/025,215
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  OSTS:002PZ1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-915-868-1


  Query Match           80.0%;  Score 14.4;  DB 4;  Length 4348;
  Best Local Similarity  93.8%;  Pred. No. 24;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAG 16
          |||||||||||||| |
Db   2219 GCCCCGATGAGCGAGG 2234


RESULT   3
US-07-798-776-1/c
; Sequence 1, Application US/07798776
; Patent No. 5434074
;  GENERAL INFORMATION:
;    APPLICANT:  GIBSON, D. WADE
;    APPLICANT:  WELCH, ANTHONY R.
;    TITLE OF INVENTION:  HERPES VIRUS PROTEINASE AND METHOD OF
;    TITLE OF INVENTION:  ASSAYING
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie & Beckett
;      STREET:  1001 G Street, N.W., Eleventh Floor
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20001-4597
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/798,776
;      FILING DATE:  19911127
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FOULKE, CYNTHIA L.
;      REGISTRATION NUMBER:  32,364
;      REFERENCE/DOCKET NUMBER:  1107.07080
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 508-9100
;      TELEFAX:  (202) 508-9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2014 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Simian cytomegalovirus
;      STRAIN:  Colburn
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  175..2001
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  175..231
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  232..2001
US-07-798-776-1


  Query Match           77.8%;  Score 14;  DB 1;  Length 2014;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGA 14
          ||||||||||||||
Db   1199 GCCCCGATGAGCGA 1186


RESULT   4
US-08-251-288A-1/c
; Sequence 1, Application US/08251288A
; Patent No. 6001967
;  GENERAL INFORMATION:
;    APPLICANT:  Gibson, Wade
;    APPLICANT:  Welch, Anthony
;    TITLE OF INVENTION:  HERPES PROTEINASE AND METHOD
;    TITLE OF INVENTION:  OF ASSAYING
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff
;      STREET:  1001 G Street, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
```

```
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/251,288A
;       FILING DATE:  31-MAY-1994
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kagan, Sarah A
;       REGISTRATION NUMBER:  32141
;       REFERENCE/DOCKET NUMBER:  01107.46284
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2014 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-251-288A-1


  Query Match             77.8%;  Score 14;  DB 5;  Length 2014;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGA 14
          ||||||||||||||
Db   1199 GCCCCGATGAGCGA 1186


RESULT   5
US-08-851-088-5
; Sequence 5, Application US/08851088
; Patent No. 5952208
;  GENERAL INFORMATION:
;    APPLICANT:  Darzins, Aldis
;    APPLICANT:  Xi, Lei
;    APPLICANT:  Childs, John D.
;    APPLICANT:  Monticello, Daniel J.
;    APPLICANT:  Squires, Charles H.
;    TITLE OF INVENTION:  DSZ Gene Expression In Pseudomonas Hosts
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.c.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/851,088
;      FILING DATE:  05-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/835,185
;      FILING DATE:  07-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmore, Carolyn S.
;      REGISTRATION NUMBER:  37,567
;      REFERENCE/DOCKET NUMBER:  EBC96-06A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (781) 861-6240
;      TELEFAX:  (781) 861-9540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1251 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1095
US-08-851-088-5


  Query Match             7( 7%;  Score 13.8;  DB 4;  Length 1251;
  Best Local Similarity   88.2%;  Pred. No. 45;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          |||||||||| |||| |
Db    301 GCCCCGATGATCGAACT 317


RESULT   6
US-08-089-755A-1
; Sequence 1, Application US/08089755A
; Patent No. 5356801
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/089,755A
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  790..2151
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  3256..4506
US-08-089-755A-1


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 5535;
  Best Local Similarity   88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          ||||||||||  |||| |
Db   3556 GCCCCGATGATCGAACT 3572


RESULT   7
US-08-089-755A-4
; Sequence 4, Application US/08089755A
; Patent No. 5356801
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/089,755A
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2148..3245
US-08-089-755A-4


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 5535;
  Best Local Similarity   88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          ||||||||||  |||| |
Db   3556 GCCCCGATGATCGAACT 3572


RESULT   8
US-08-421-754-1
; Sequence 1, Application US/08421754
; Patent No. 5578478
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/421,754
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/304,081
;      FILING DATE:  01-SEP-1994
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/089,755
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03AZX
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  790..2151
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3256..4506
US-08-421-754-1


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 5535;
  Best Local Similarity   88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGT 17
          ||||||||||  |||| |
Db   3556 GCCCCGATGATCGAACT 3572
```

```
RESULT  9
US-08-421-754-4
; Sequence 4, Application US/08421754
; Patent No. 5578478
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/421,754
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/304,081
;      FILING DATE:  01-SEP-1994
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/089,755
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03AZX
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2148..3245
US-08-421-754-4


  Query Match             76.7%;  Score 13.8;  DB 1;  Length 5535;
  Best Local Similarity   88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGT 17
           ||||||||||| |||| |
Db    3556 GCCCCGATGATCGAACT 3572


RESULT  10
US-08-421-791-1
; Sequence 1, Application US/08421791
; Patent No. 5879914
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/421,791
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/304,081
;      FILING DATE:  01-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/089,755
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03AZ2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  790..2151
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3256..4506
US-08-421-791-1


  Query Match             76.7%;  Score 13.8;  DB 3;  Length 5535;
  Best Local Similarity   88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGT 17
           ||||||||||| |||| |
Db    3556 GCCCCGATGATCGAACT 3572


RESULT  11
US-08-421-791-4
; Sequence 4, Application US/08421791
; Patent No. 5879914
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Rambosek, John
;    APPLICANT:  Piddington, Chris
;    APPLICANT:  Kovacevich, Brian R
;    APPLICANT:  Young, Kevin D
;    APPLICANT:  Denome, Sylvia A
;    TITLE OF INVENTION:  Recombinant DNA Encoding A
;    TITLE OF INVENTION:  Desulfurization Biocatalyst
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith and Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/421,791
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/304,081
;      FILING DATE:  01-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/089,755
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/911,845
;      FILING DATE:  10-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  EBC92-03AZ2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2148..3245
US-08-421-791-4


  Query Match           76.7%;  Score 13.8;  DB 3;  Length 5535;
  Best Local Similarity  88.2%;  Pred. No. 52;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGT 17
           ||||||||||  |||| |
Db    3556 GCCCCGATGATCGAACT 3572


RESULT  12
US-08-846-111D-15/c
; Sequence 15, Application US/08846111D
; Patent No. 6017705
;  GENERAL INFORMATION:
;    APPLICANT:  Lurquin, Christophe; Brasseur, Francis;
;    APPLICANT:  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Which Are
;    TITLE OF INVENTION:  Members Of The Mage-B Family and Uses Thereof
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect For DOS 6.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,111D
;      FILING DATE:  25-APRIL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/658,578
;      FILING DATE:  5-MAY-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/403,388
;      FILING DATE:  14-MARCH-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6017705man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5444.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40352 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-846-111D-15


  Query Match           76.7%;  Score 13.8;  DB 5;  Length 40352;
  Best Local Similarity  88.2%;  Pred. No. 63;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 CCCCGATGAGCGAAGTC 18
           ||| |||||| ||||||
Db   29386 CCCAGATGAGGGAAGTC 29370


RESULT  13
US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 5025546
; GENERAL INFORMATION:
;  APPLICANT: PLANT GENETIC SYSTEMS N.V.
;  TITLE OF INVENTION: Method to obtain male sterile plants
;  FILE REFERENCE: NMSCOR
;  CURRENT APPLICATION NUMBER: US/08/894,440
;  CURRENT FILING DATE: 1997-11-12
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1303
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
;   OTHER INFORMATION: fragment of pTS88
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(35)
;   OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
;   FEATURE:
;   NAME/KEY: misc_feature
```

; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
US-08-894-440-2

Query Match 74.4%; Score 13.4; DB 5; Length 1303;
Best Local Similarity 93.3%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCCGATGAGCGAAG 16
||||||||||| |||
Db 1029 ccccgatgagctaag 1043

RESULT 14
US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maike
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-453-104-23

Query Match 74.4%; Score 13.4; DB 1; Length 3200;
Best Local Similarity 93.3%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CCCCGATGAGCGAAG 16
||||||||||| |||
Db 2883 CCCCGATGAGCTAAG 2897

RESULT 15
US-08-694-824-23
; Sequence 23, Application US/08694824
; Patent No. 5877306
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maike
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082

; OTHER INFORMATION: wherein N is not known."
US-08-694-824-23

 Query Match 74.4%; Score 13.4; DB 3; Length 3200;
 Best Local Similarity 93.3%; Pred. No. 82;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       2 CCCCGATGAGCGAAG 16
           ||||||||||| |||
Db    2883 CCCCGATGAGCTAAG 2897
```

Search completed: June 22, 2000, 15:21:57
Job time: 18795 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:22 ; Search time 5541.94 Seconds
(without alignments)
13.165 Million cell updates/sec

Title: US-09-362-485-15
Perfect score: 18
Sequence: 1 GCCCCGATGAGCGAAGTC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	15.4	85.6	491	105	AQ303458	AQ303458 HS_3227_A
c	2	15	83.3	676	94	AQ074000	AQ074000 EP(3)3424
c	3	14.8	82.2	77	42	AI158677	AI158677 ud29h11.r
	4	14.8	82.2	449	28	AA119021	AA119021 mo28d03.r
c	5	14.8	82.2	489	46	AI448289	AI448289 mo48c07.x
c	6	14.8	82.2	499	103	AQ205008	AQ205008 HS_3231_A
c	7	14.8	82.2	516	29	AA155083	AA155083 ms40h01.r
c	8	14.8	82.2	516	29	AA164004	AA164004 ms40h09.r
c	9	14.8	82.2	613	35	AA552378	AA552378 nk13f02.s
c	10	14.8	82.2	670	45	AI386158	AI386158 mq49f05.y
	11	14.8	82.2	683	40	AA951313	AA951313 LD31719.5
c	12	14.8	82.2	727	45	AI393497	AI393497 tf77e01.x
c	13	14.8	82.2	1101	82	CNS0023A	AL097168 Drosophil
	14	14.4	80.0	137	80	AW374931	AW374931 MR0-CT006
	15	14.4	80.0	289	72	AV325088	AV325088 AV325088
c	16	14.4	80.0	305	35	AA590445	AA590445 vm20c11.r
	17	14.4	80.0	308	60	AV147273	AV147273 AV147273
	18	14.4	80.0	323	69	AW124157	AW124157 UI-M-BH2.
c	19	14.4	80.0	345	82	FR0024096	AL016947 F.rubripe
c	20	14.4	80.0	355	48	AI604536	AI604536 vm20c11.y
	21	14.4	80.0	357	47	AI503823	AI503823 vm20c11.x
c	22	14.4	80.0	469	82	FR0024222	AL017073 F.rubripe
	23	14.4	80.0	548	82	FR0024244	AL017095 F.rubripe
c	24	14.4	80.0	567	29	AA164204	AA164204 zq46f02.s
c	25	14.4	80.0	582	30	AA206097	AA206097 zq54f04.s
c	26	14.4	80.0	590	82	FR0024240	AL017091 F.rubripe
	27	14.4	80.0	641	80	AW392282	AW392282 QV0-ST023
c	28	14	77.8	296	91	AQ847179	AQ847179 LMAJFV1_1
c	29	14	77.8	479	82	CNS00UMG	AL091174 Arabidops
	30	13.8	76.7	224	20	D39521	D39521 RICS0924A R
c	31	13.8	76.7	230	23	H50794	H50794 yp10d02.r1
c	32	13.8	76.7	241	72	AV288400	AV288400 AV288400
c	33	13.8	76.7	253	22	R51867	R51867 yg74g05.s1
	34	13.8	76.7	255	31	AA326263	AA326263 EST29426
c	35	13.8	76.7	271	80	AW346474	AW346474 28167 MAR
c	36	13.8	76.7	275	72	AV329539	AV329539 AV329539
	37	13.8	76.7	291	61	AV164343	AV164343 AV164343
	38	13.8	76.7	294	109	AQ657609	AQ657609 Sheared D
c	39	13.8	76.7	304	80	AW326412	AW326412 18984 MAR
	40	13.8	76.7	323	40	AA987594	AA987594 or84h09.s
	41	13.8	76.7	329	50	AI698830	AI698830 wd31e03.x
c	42	13.8	76.7	330	26	W79036	W79036 zh50a04.r1
c	43	13.8	76.7	336	24	H67060	H67060 yu68f01.r1
c	44	13.8	76.7	338	34	AA481385	AA481385 zv44h06.r
c	45	13.8	76.7	338	81	AW437098	AW437098 77895 MAR

ALIGNMENTS

```
RESULT   1
AQ303458/c
LOCUS       AQ303458      491 bp    DNA             GSS       16-DEC-1998
DEFINITION  HS_3227_A2_B11_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3227 Col=22 Row=C, genomic survey
            sequence.
ACCESSION   AQ303458
VERSION     AQ303458.1  GI:4023685
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-7618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3227  row: C  column: 22
            Class: BAC ends
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3227 Col=22 Row=C"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      121 a    143 c     93 g    126 t      8 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 105;  Length 491;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGT 17
           ||||||||||| |||||
Db     400 GCCCCGATGAGAGAAGT 384


RESULT   2
AQ074000/c
LOCUS       AQ074000      676 bp    DNA             GSS       28-JUN-1999
DEFINITION  EP(3)3424 Drosophila melanogaster EP line Drosophila melanogaster
            genomic Sequence recovered from Both 5' and 3' ends of P element,
            genomic survey sequence.
ACCESSION   AQ074000
VERSION     AQ074000.1  GI:3404042
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Rehm,E.J. and Rubin,G.M.
  TITLE     The BDGP gene disruption project: single EP element insertions
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu

            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 548 in the 676 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.

            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster EP line"
```

/note="Inverse PCR was performed on Drosophila melanogaster strains each of which contains a single EP transposable element insertion. (The generation of these insertion strains is described in Rorth P, Szabo K, Bailey A, Laverty T, Rehm J, Rubin GM, Weigmann K, Milan M, Benes V, Ansorge W, Cohen SM. 1998. Systematic gain-of-function genetics in Drosophila. Development 6:1049-1057.) The resultant fragment for each strain was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT 167 a 164 c 153 g 191 t 1 others
ORIGIN

Query Match 83.3%; Score 15; DB 94; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 CCGATGAGCGAAGTC 18
||||||||||||||||
Db 615 CCGATGAGCGAAGTC 601

RESULT 3
AI158677/c
LOCUS AI158677 77 bp mRNA EST 30-SEP-1998
DEFINITION ud29h11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1447365 5' similar to gb:M97957 Mouse pyruvate carboxylase homologous protein mRNA, 3' (MOUSE);, mRNA sequence.
ACCESSION AI158677
VERSION AI158677.1 GI:3687146
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 77)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285459.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:920681
Trace considered overall poor quality
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .77
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1447365"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 17 a 15 c 29 g 16 t
ORIGIN

Query Match 82.2%; Score 14.8; DB 42; Length 77;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCCCCGATGAGCGAAGTC 18
||| | ||||||||||||
Db 47 GCCTCCATGAGCGAAGTC 30

RESULT 4
AA119021
LOCUS AA119021 449 bp mRNA EST 19-NOV-1996
DEFINITION mo28d03.r1 Life Tech mouse embryo 13 5dpc 10666014 Mus musculus cDNA clone IMAGE:554885 5' similar to gb:X12451 CATHEPSIN L PRECURSOR (HUMAN); gb:X06086 Mouse mRNA for major excreted protein (MOUSE);, mRNA sequence.
ACCESSION AA119021
VERSION AA119021.1 GI:1676663
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 449)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407053.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:335677
Seq primer: -28M13 rev1 from Amersham
High quality sequence stop: 127.
FEATURES Location/Qualifiers
source 1. .449
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:554885"
/clone_lib="Life Tech mouse embryo 13 5dpc 10666014"
/tissue_type="embryo"
/dev_stage="13.5dpc embryos"
/lab_host="DH10B"
/note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. 13.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT 115 a 90 c 149 g 95 t
ORIGIN

```
  Query Match            82.2%;  Score 14.8;  DB 28;  Length 449;
  Best Local Similarity  88.9%;  Pred. No. 2.7e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||| |||||||||| |
Db    362 GCCCCTATGAGCGAAGAC 379


RESULT   5
AI448289/c
LOCUS       AI448289      489 bp    mRNA            EST       09-MAR-1999
DEFINITION  mo48c07.x1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus
            cDNA clone IMAGE:556812 3', mRNA sequence.
ACCESSION   AI448289
VERSION     AI448289.1  GI:4281446
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187736.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:556812"
                     /clone_lib="Life Tech mouse embryo 10 5dpc 10665016"
                     /tissue_type="embryo"
                     /dev_stage="10.5dpc embryos"
                     /lab_host="DH10B"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  10.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT      144 a    113 c    110 g    122 t
ORIGIN


  Query Match            82.2%;  Score 14.8;  DB 46;  Length 489;
  Best Local Similarity  88.9%;  Pred. No. 2.8e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          ||||| |||||||||| |
Db    484 GCCCCGCTGAGCAAAGTC 467


RESULT   6
AQ205008/c
LOCUS       AQ205008      499 bp    DNA             GSS       17-SEP-1998
DEFINITION  HS_3231_A2_H06_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3231 Col=12 Row=O, genomic survey
            sequence.
ACCESSION   AQ205008
VERSION     AQ205008.1  GI:3615578
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3231  row: O  column: 12
            Class: BAC ends
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3231 Col=12 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      153 a     87 c     95 g    153 t     11 others
ORIGIN


  Query Match            82.2%;  Score 14.8;  DB 103;  Length 499;
  Best Local Similarity  88.9%;  Pred. No. 2.8e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GCCCCGATGAGCGAAGTC 18
          |||| ||||||| |||||
Db    481 GCCCAGATGAGCCAAGTC 464


RESULT   7
AA155083/c
LOCUS       AA155083      516 bp    mRNA            EST       11-DEC-1996
DEFINITION  ms40h01.r1 Life Tech mouse embryo 13 5dpc 10666014 Mus musculus
            cDNA clone IMAGE:614065 5' similar to gb:M17885 60S ACIDIC
            RIBOSOMAL PROTEIN P0 (HUMAN); gb:X15267 Mouse mRNA for acidic
            ribosomal phosophoprotein PO (MOUSE);, mRNA sequence.
ACCESSION   AA155083
VERSION     AA155083.1  GI:1726899
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1326658.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
```

```
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:374889
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 158.
FEATURES              Location/Qualifiers
     source           1. .516
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:614065"
                      /clone_lib="Life Tech mouse embryo 13 5dpc 10666014"
                      /tissue_type="embryo"
                      /dev_stage="13.5dpc embryos"
                      /lab_host="DH10B"
                      /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                      SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                      Oligo dT.  13.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT      127 a    133 c    151 g    105 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 29;  Length 516;
  Best Local Similarity   88.9%;  Pred. No. 2.8e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           |||  ||||||||||||||
Db     459 GCCTTGATGAGCGAAGTC 442


RESULT   8
AA164004/c
LOCUS       AA164004      516 bp    mRNA            EST       17-DEC-1996
DEFINITION  ms40h09.r1 Life Tech mouse embryo 13 5dpc 10666014 Mus musculus
            cDNA clone IMAGE:614081 5' similar to gb:M17885 60S ACIDIC
            RIBOSOMAL PROTEIN P0 (HUMAN); gb:X15267 Mouse mRNA for acidic
            ribosomal phosophoprotein P0 (MOUSE);, mRNA sequence.
ACCESSION   AA164004
VERSION     AA164004.1  GI:1739966
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Marra,M., Fillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637809.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:374905
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 158.
FEATURES              Location/Qualifiers
     source           1. .516
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:614081"
                      /clone_lib="Life Tech mouse embryo 13 5dpc 10666014"
                      /tissue_type="embryo"
                      /dev_stage="13.5dpc embryos"
                      /lab_host="DH10B"
                      /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                      SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                      Oligo dT.  13.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT      127 a    133 c    151 g    105 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 29;  Length 516;
  Best Local Similarity   88.9%;  Pred. No. 2.8e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           |||  ||||||||||||||
Db     459 GCCTTGATGAGCGAAGTC 442


RESULT   9
AA552378/c
LOCUS       AA552378      613 bp    mRNA            EST       05-SEP-1997
DEFINITION  nk13f02.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1013403 3'
            similar to gb:J04173 PHOSPHOGLYCERATE MUTASE, BRAIN FORM (HUMAN);,
            mRNA sequence.
ACCESSION   AA552378
VERSION     AA552378.1  GI:2322630
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
            Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 703    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 381.
FEATURES              Location/Qualifiers
     source           1. .613
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1013403"
                      /clone_lib="NCI_CGAP_Co2"
                      /tissue_type="tumor"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: colon; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
                      5' GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3' Average insert size:  1.1 kb."
BASE COUNT      186 a    135 c    137 g    155 t
ORIGIN
```

```
  Query Match             82.2%;  Score 14.8;  DB 35;  Length 613;
  Best Local Similarity   88.9%;  Pred. No. 3e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           | ||| ||||||||||||
Db     557 GTCCCCATGAGCGAAGTC 540


RESULT  10
AI386158/c
LOCUS       AI386158      670 bp    mRNA            EST       27-JAN-1999
DEFINITION  mq49f05.y1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:582081
            5' similar to gb:X68277_cds1 MAP KINASE PHOSPHATASE-1 (HUMAN);
            gb:X61940 Mouse mRNA for a growth factor-inducible immediate early
            gene (MOUSE);, mRNA sequence.
ACCESSION   AI386158
VERSION     AI386158.1  GI:4199621
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 15, 1998 this sequence version replaced gi:3224923.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:356729
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1. .670
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:582081"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      144 a    202 c    164 g    157 t      3 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 45;  Length 670;
  Best Local Similarity   88.9%;  Pred. No. 3.1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           | ||||||| ||||||||
Db     588 GTCCCGATGTGCGAAGTC 571


RESULT  11
AA951313
LOCUS       AA951313      683 bp    mRNA            EST       25-NOV-1998
DEFINITION  LD31719.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD31719 5prime, mRNA sequence.
ACCESSION   AA951313
VERSION     AA951313.1  GI:3113550
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404838.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 317  row: B  column: 7
            High quality sequence stop: 517.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD31719"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      173 a    175 c    187 g    148 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 40;  Length 683;
  Best Local Similarity   88.9%;  Pred. No. 3.1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           |||||||||||||| | |
Db     300 GCCCCGATGAGCGACGAC 317


RESULT  12
AI393497/c
LOCUS       AI393497      727 bp    mRNA            EST       30-MAR-1999
DEFINITION  tf77e01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2105304 3'
            similar to gb:J04173 PHOSPHOGLYCERATE MUTASE, BRAIN FORM (HUMAN);,
            mRNA sequence.
ACCESSION   AI393497
VERSION     AI393497.1  GI:4223044
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
```

```
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2153539.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1229   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2105304"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      223 a    164 c    158 g    182 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 45;  Length 727;
  Best Local Similarity   88.9%;  Pred. No. 3.2e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           | ||| ||||||||||||
Db     558 GTCCCCATGAGCGAAGTC 541


RESULT  13
CNS00Z3A/c
LOCUS       CNS00Z3A     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01C02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097168
VERSION     AL097168.1  GI:5608779
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN01C02"
                     /note="end : T7"
BASE COUNT      262 a    233 c    196 g    253 t    157 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 82;  Length 1101;
  Best Local Similarity   88.9%;  Pred. No. 3.7e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GCCCCGATGAGCGAAGTC 18
           |||||||||||||| | |
Db     463 GCCCCGATGAGCGACGAC 446


RESULT  14
AW374931
LOCUS       AW374931      137 bp    mRNA            EST       04-FEB-2000
DEFINITION  MR0-CT0062-151099-004-b12 CT0062 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW374931
VERSION     AW374931.1  GI:6879585
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6676857.
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-CT0062-
            151099-004-b12&t3=1999-10-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 78.
FEATURES             Location/Qualifiers
     source          1. .137
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0062"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: pUC18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No.
                     196,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       37 a     38 c     38 g     24 t
ORIGIN


  Query Match             80.0%;  Score 14.4;  DB 80;  Length 137;
  Best Local Similarity   93.8%;  Pred. No. 3e+02;
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-033-18


  Query Match            82.2%;  Score 14.8;  DB 2;  Length 2004;
  Best Local Similarity  88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          |||||||||| || ||||
Db    844 GGCGGCGAGTGGAACGGC 861


RESULT   5
US-08-471-044-18
; Sequence 18, Application US/08471044
```

```
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-044-18


  Query Match            82.2%;  Score 14.8;  DB 3;  Length 2004;
  Best Local Similarity  88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          |||||||||| || ||||
Db    844 GGCGGCGAGTGGAACGGC 861
```

```
RESULT   6
US-08-463-483A-18
; Sequence 18, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-463-483A-18


  Query Match           82.2%;  Score 14.8;  DB 3;  Length 2004;
  Best Local Similarity  88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db     844 GGCGGCGAGTGGAACGGC 861
```

```
RESULT   7
US-08-471-046A-18
; Sequence 18, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-199
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-046A-18


  Query Match           82.2%;  Score 14.8;  DB 3;  Length 2004;
  Best Local Similarity  88.9%;  Pred. No. 56;
```

```
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db     844 GGCGGCGAGTGGAACGGC 861


RESULT   8
US-08-470-566B-18
; Sequence 18, Application US/08470566B
; Patent No. 5872212
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5872212el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5872212artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,566B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV4 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-470-566B-18

  Query Match           82.2%;  Score 14.8;  DB 3;  Length 2004;
  Best Local Similarity   88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db     844 GGCGGCGAGTGGAACGGC 861


RESULT   9
US-08-469-334-18
; Sequence 18, Application US/08469334
; Patent No. 5990383
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5990383el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,334
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/463,483
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2004 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2004
```

```
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for VIP1A(a) 80 kd protein from AB78"
US-08-469-334-18


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 2004;
  Best Local Similarity   88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||  || ||||
Db     844 GGCGGCGAGTGGAACGGC 861


RESULT  10
US-08-471-033-35
; Sequence 35, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..2564
;      OTHER INFORMATION:  /note= "Maize optimized sequence
;      OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal remov
;      OTHER INFORMATION:  contained in pCIB5526"
US-08-471-033-35


  Query Match             82.2%;  Score 14.8;  DB 2;  Length 2576;
  Best Local Similarity   88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||  || ||||
Db    1407 GGCGGCGAGTGGAACGGC 1424


RESULT  11
US-08-471-044-35
; Sequence 35, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..2564
;      OTHER INFORMATION:  /note= "Maize optimized sequence
;      OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal removed
;      OTHER INFORMATION:  contained in pCIB5526"
US-08-471-044-35


  Query Match             82.2%;  Score 14.8;  DB 3;  Length 2576;
  Best Local Similarity   88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db    1407 GGCGGCGAGTGGAACGGC 1424


RESULT  12
US-08-463-483A-35
; Sequence 35, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..2564
;      OTHER INFORMATION:  /note= "Maize optimized sequence
;      OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal remov
;      OTHER INFORMATION:  contained in pCIB5526"
US-08-463-483A-35


  Query Match             82.2%;  Score 14.8;  DB 3;  Length 2576;
  Best Local Similarity   88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db    1407 GGCGGCGAGTGGAACGGC 1424


RESULT  13
US-08-471-046A-35
; Sequence 35, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  35:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2576 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "Synthetic DNA"
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  9..2564
;       OTHER INFORMATION:  /note= "Maize optimized sequence
;       OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal removed
;       OTHER INFORMATION:  contained in pCIB5526"
US-08-471-046A-35


  Query Match           82.2%;  Score 14.8;  DB 3;  Length 2576;
  Best Local Similarity  88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||  || ||||
Db   1407 GGCGGCGAGTGGAACGGC 1424


RESULT  14
US-08-470-566B-35
; Sequence 35, Application US/08470566B
; Patent No. 5872212
;  GENERAL INFORMATION:
;     APPLICANT:  Warren, Gregory W
;     APPLICANT:  Koziel, Michael G
;     APPLICANT:  Mullins, Martha A
;     APPLICANT:  Nye, Gordon J
;     APPLICANT:  Carr, Brian
;     APPLICANT:  Desai, Nalini M
;     APPLICANT:  Kostichka, N. Kristy
;     APPLICANT:  Duck, Nicholas B
;     APPLICANT:  Estruch, Juan J
;     TITLE OF INVENTION:  No. 5872212el Pesticidal Proteins and Strains
;     NUMBER OF SEQUENCES:  52
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  No. 5872212artis Corporation
;       STREET:  3054 Cornwallis Road
;       CITY:  Research Triangle Park
;       STATE:  NC
;       COUNTRY:  USA
;       ZIP:  27709
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/470,566B
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/463,483
;       FILING DATE:  05-JUN-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/314,594
;       FILING DATE:  09-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/218,018
;       FILING DATE:  23-MAR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/037,057
;       FILING DATE:  25-MAR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meigs, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DLV4 - SQLv4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8587
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  35:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2576 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "Synthetic DNA"
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  9..2564
;       OTHER INFORMATION:  /note= "Maize optimized sequence
;       OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal remov
;       OTHER INFORMATION:  contained in pCIB5526"
US-08-470-566B-35


  Query Match           82.2%;  Score 14.8;  DB 3;  Length 2576;
  Best Local Similarity  88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||  || ||||
Db   1407 GGCGGCGAGTGGAACGGC 1424


RESULT  15
US-08-469-334-35
; Sequence 35, Application US/08469334
; Patent No. 5990383
;  GENERAL INFORMATION:
;     APPLICANT:  Warren, Gregory W
;     APPLICANT:  Koziel, Michael G
;     APPLICANT:  Mullins, Martha A
;     APPLICANT:  Nye, Gordon J
;     APPLICANT:  Carr, Brian
;     APPLICANT:  Desai, Nalini M
;     APPLICANT:  Kostichka, N. Kristy
;     APPLICANT:  Duck, Nicholas B
;     APPLICANT:  Estruch, Juan J
;     TITLE OF INVENTION:  No. 5990383el Pesticidal Proteins and Strains
;     NUMBER OF SEQUENCES:  50
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CIBA-GEIGY Corporation
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,334
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/463,483
;       FILING DATE:
;       APPLICATION NUMBER:  US 08/314,594
;       FILING DATE:  09-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/218,018
```

```
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2576 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..2564
;      OTHER INFORMATION:  /note= "Maize optimized sequence
;      OTHER INFORMATION:  encoding VIP1A(a) with the Bacillus secretion signal removed
;      OTHER INFORMATION:  contained in pCIB5526"
US-08-469-334-35


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 2576;
  Best Local Similarity   88.9%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| || ||||
Db    1407 GGCGGCGAGTGGAACGGC 1424


Search completed: June 22, 2000, 18:04:11
Job time: 9718 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:42 ; Search time 2113.77 Seconds
(without alignments)
34.516 Million cell updates/sec

Title: US-09-362-485-25
Perfect score: 18
Sequence: 1 GGCGGCGAGTCGACCGGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	15.4	85.6	341	81	AJ285218	AJ285218 4A3B-AAH-
	2	15.4	85.6	352	81	AJ284714	AJ284714 4A3B-AAB-
	3	15.4	85.6	406	81	AJ284812	AJ284812 4A3B-AAC-
c	4	15.4	85.6	418	64	AW099909	AW099909 sd18c04.y
c	5	15.4	85.6	439	81	AJ283520	AJ283520 4A3B-AAC-
c	6	14.8	82.2	140	28	AA104391	AA104391 mo42b04.r
	7	14.8	82.2	203	36	C71766	C71766 C71766 Rice
c	8	14.8	82.2	267	33	AA409128	AA409128 EST03607
c	9	14.8	82.2	319	37	AA685482	AA685482 EST107297
c	10	14.8	82.2	333	42	AI131683	AI131683 ue95e06.x
c	11	14.8	82.2	345	25	W09281	W09281 ma06b04.r1
c	12	14.8	82.2	360	26	W91633	W91633 MTA.H10.085
c	13	14.8	82.2	360	74	AW229945	AW229945 uo45b04.y
	14	14.8	82.2	363	43	AU032296	AU032296 AU032296
	15	14.8	82.2	380	48	AI625890	AI625890 CrESTz001
	16	14.8	82.2	390	49	AI662864	AI662864 CrESTz005
c	17	14.8	82.2	395	25	W11142	W11142 ma74b02.r1
c	18	14.8	82.2	404	79	AL024098	AL024098 r8224b45
c	19	14.8	82.2	422	94	AQ074265	AQ074265 81 pUC8 P
c	20	14.8	82.2	426	37	AA691834	AA691834 vt05d09.r
c	21	14.8	82.2	435	50	AI677540	AI677540 60056G02
c	22	14.8	82.2	444	35	AA560311	AA560311 vl67c11.r
c	23	14.8	82.2	451	27	W16262	W16262 mb64c07.r1
	24	14.8	82.2	454	35	C27017	C27017 C27017 Rice
	25	14.8	82.2	459	64	AW099947	AW099947 sd18g04.y
c	26	14.8	82.2	460	25	W14494	W14494 mb39g02.r1
	27	14.8	82.2	463	74	AV397327	AV397327 AV397327
c	28	14.8	82.2	465	28	AA102945	AA102945 mo09b02.r
c	29	14.8	82.2	467	36	AA638343	AA638343 vl98f01.r
	30	14.8	82.2	467	62	AU077787	AU077787 AU077787
c	31	14.8	82.2	468	30	AA277644	AA277644 va80d01.r
c	32	14.8	82.2	470	37	AA692088	AA692088 vt19a06.r
	33	14.8	82.2	475	62	AU077786	AU077786 AU077786
c	34	14.8	82.2	476	44	AI286510	AI286510 ui84h10.y
c	35	14.8	82.2	483	92	AQ911436	AQ911436 LMAJFV1_1
c	36	14.8	82.2	498	38	AA796376	AA796376 vs99d11.r
c	37	14.8	82.2	510	33	AA387268	AA387268 vc24c10.r
c	38	14.8	82.2	511	35	AA592168	AA592168 vo19f10.r
c	39	14.8	82.2	515	27	AA002898	AA002898 mg41h12.r
c	40	14.8	82.2	515	48	AI574341	AI574341 uk19h01.y
c	41	14.8	82.2	518	62	AI931242	AI931242 ul58f08.y
c	42	14.8	82.2	519	42	AI119029	AI119029 ue22h08.y
c	43	14.8	82.2	524	62	AI901947	AI901947 618012H04
c	44	14.8	82.2	528	27	AA023063	AA023063 mh66b04.r
c	45	14.8	82.2	533	79	AL023048	AL023048 r8220b26

ALIGNMENTS

```
RESULT   1
AJ285218
LOCUS       AJ285218      341 bp    mRNA            EST       07-FEB-2000
DEFINITION  4A3B-AAH-A-08-R Anopheles gambiae immune competent 4A3B Anopheles
            gambiae cDNA clone 4A3B-AAH-A-08, mRNA sequence.
ACCESSION   AJ285218
VERSION     AJ285218.1  GI:6933099
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae pilot gene discovery project: Identification of
            novel mosquito innate immunity genes from ESTs generated from
            immune competent cell lines
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6676768.
            Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Anopheles gambiae"
                     /strain="4A r/r"
                     /db_xref="taxon:7165"
                     /clone="4A3B-AAH-A-08"
                     /clone_lib="Anopheles gambiae immune competent 4A3B"
                     /cell_line="immune competent 4A3B"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from
                     the forward priming site that reads from the 3' end of
                     cDNA. The 4A3B is a directionally cloned and normalized
                     cDNA library that was constructed from the 4A3B cell line
                     oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
                     (1996): Normalization and Subtraction: Two Approaches To
                     Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT       81 a    108 c     96 g     56 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 81;  Length 341;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGG 17
           ||||||| ||||||||||
Db     190 GGCGGCA\GTCGACCGG 206


RESULT   2
AJ284714
LOCUS       AJ284714      352 bp    mRNA            EST       07-FEB-2000
DEFINITION  4A3B-AAB-E-01-R Anopheles gambiae immune competent 4A3B Anopheles
            gambiae cDNA clone 4A3B-AAB-E-01, mRNA sequence.
ACCESSION   AJ284714
VERSION     AJ284714.1  GI:6932593
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae pilot gene discovery project: Identification of
            novel mosquito innate immunity genes from ESTs generated from
            immune competent cell lines
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6675739.
            Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Anopheles gambiae"
                     /strain="4A r/r"
                     /db_xref="taxon:7165"
                     /clone="4A3B-AAB-E-01"
                     /clone_lib="Anopheles gambiae immune competent 4A3B"
                     /cell_line="immune competent 4A3B"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from
                     the forward priming site that reads from the 3' end of
```

cDNA. The 4A3B is a directionally cloned and normalized cDNA library that was constructed from the 4A3B cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT 83 a 110 c 104 g 55 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 81; Length 352;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGG 17
|||||| ||||||||||
Db 181 GGCGGCAAGTCGACCGG 197

RESULT 3
AJ284812
LOCUS AJ284812 406 bp mRNA EST 07-FEB-2000
DEFINITION 4A3B-AAC-F-01-R Anopheles gambiae immune competent 4A3B Anopheles gambiae cDNA clone 4A3B-AAC-F-01, mRNA sequence.
ACCESSION AJ284812
VERSION AJ284812.1 GI:6932691
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anopheles.
REFERENCE 1 (bases 1 to 406)
AUTHORS Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.
TITLE Anopheles gambiae pilot gene discovery project: Identification of novel mosquito innate immunity genes from ESTs generated from immune competent cell lines
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6675840.
Contact: Dimopoulos G
Fotis C. Kafatos laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES Location/Qualifiers
source 1. .406
/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3B-AAC-F-01"
/clone_lib="Anopheles gambiae immune competent 4A3B"
/cell_line="immune competent 4A3B"
/lab_host="E. coli DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from the forward priming site that reads from the 3' end of cDNA. The 4A3B is a directionally cloned and normalized cDNA library that was constructed from the 4A3B cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT 97 a 123 c 113 g 73 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 81; Length 406;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGG 17
|||||| ||||||||||
Db 254 GGCGGCAAGTCGACCGG 270

RESULT 4
AW099909/c
LOCUS AW099909 418 bp mRNA EST 06-DEC-1999
DEFINITION sd18c04.y2 Gm-c1012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1012-2047 5', mRNA sequence.
ACCESSION AW099909
VERSION AW099909.1 GI:6070522
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 418)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3137759.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1. .418
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1012-2047"
/clone_lib="Gm-c1012"
/tissue_type="Apical shoot tips, 9-10 day old etiolated seedlings"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from the apical shoots of 9 to 10 day old etiolated seedlings. The shoot tips including any emerged leaves were harvested for mRNA isolation. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."
BASE COUNT 95 a 104 c 137 g 82 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 64; Length 418;
Best Local Similarity 94.1%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGG 17

```
              |||||||||||| ||||
Db    270 GGCGGCGAGTCGTCCGG 254


RESULT   5
AJ283520/c
LOCUS       AJ283520      439 bp    mRNA            EST       07-FEB-2000
DEFINITION  4A3B-AAC-F-01-F Anopheles gambiae immune competent 4A3B Anopheles
            gambiae cDNA clone 4A3B-AAC-F-01, mRNA sequence.
ACCESSION   AJ283520
VERSION     AJ283520.1  GI:6931399
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
  TITLE     Anopheles gambiae pilot gene discovery project: Identification of
            novel mosquito innate immunity genes from ESTs generated from
            immune competent cell lines
  JOURNAL   Unpublished (2000)
COMMENT     On Jul 9, 1999 this sequence version replaced gi:5435187.
            Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Anopheles gambiae"
                     /strain="4A r/r"
                     /db_xref="taxon:7165"
                     /clone="4A3B-AAC-F-01"
                     /clone_lib="Anopheles gambiae immune competent 4A3B"
                     /cell_line="immune competent 4A3B"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from
                     the forward priming site that reads from the 3' end of
                     cDNA. The 4A3B is a directionally cloned and normalized
                     cDNA library that was constructed from the 4A3B cell line
                     oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
                     (1996): Normalization and Subtraction: Two Approaches To
                     Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT       79 a    125 c    106 g    129 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 81;  Length 439;
  Best Local Similarity   94.1%;  Pred. No. 1.3e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGG 17
          |||||| ||||||||||
Db    423 GGCGGCAAGTCGACCGG 407


RESULT   6
AA104391/c
LOCUS       AA104391      140 bp    mRNA            EST       29-OCT-1996
DEFINITION  mo42b04.r1 Life Tech mouse embryo 15 5dpc 10667012 Mus musculus
            cDNA clone IMAGE:556207 5' similar to gb:Z28407 60S RIBOSOMAL
            PROTEIN L8 (HUMAN);, mRNA sequence.
ACCESSION   AA104391
VERSION     AA104391.1  GI:1650568
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 140)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637727.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:336999
            Seq primer: -28M13 rev1 from Amersham
            High quality sequence stop: 93.
FEATURES             Location/Qualifiers
     source          1. .140
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:556207"
                     /clone_lib="Life Tech mouse embryo 15 5dpc 10667012"
                     /tissue_type="embryo"
                     /dev_stage="15.5dpc embryos"
                     /lab_host="DH10B"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT.  15.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT       22 a     40 c     51 g     27 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 28;  Length 140;
  Best Local Similarity   88.9%;  Pred. No. 2e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          |||||||||||  |||||
Db     25 GGCGGCGAGTCTGCCGGC 8


RESULT   7
C71766
LOCUS       C71766        203 bp    mRNA            EST       22-SEP-1997
DEFINITION  C71766 Rice panicle at flowering stage Oryza sativa cDNA clone
            E0316_1A, mRNA sequence.
ACCESSION   C71766
VERSION     C71766.1  GI:2427303
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 203)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle at flowering stage
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400826.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
```

```
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .203
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E0316_1A"
                     /clone_lib="Rice panicle at flowering stage"
                     /dev_stage="flowering stage"
                     /note="Organ: panicle; Rice cDNA from panicle at flowering
                     stage"
BASE COUNT       27 a      81 c      66 g      26 t       3 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 36;  Length 203;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| ||||| ||
Db     138 GGCGGCGAGGCGACCTGC 155


RESULT   8
AA409128/c
LOCUS       AA409128      267 bp    mRNA            EST       26-AUG-1998
DEFINITION  EST03607 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus
            musculus cDNA clone C0037E03 5', mRNA sequence.
ACCESSION   AA409128
VERSION     AA409128.1  GI:2068522
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Ko,M.S.H., Threat,T.A., Horton,J.H., Wang,X, Cui Y., Wang,X.,
            Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and
            Nakashima,H.
  TITLE     Systematic analyses of mouse genes expressed in embryo implantation
            site
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394861.
            Other_ESTs: EST03606
            Contact: Ko MSH
            Center for Molecular Medicine and Genetics
            Wayne State University
            5047 Gullen Mall, Detroit, MI 48202
            Tel: 3135776708
            Fax: 3135776200
            Email: msko@cmb.biosci.wayne.edu
            Mus musculus ribosomal protein L8 (RPL8) mRNA, complete cds
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="ATCC (inhost):1364169"
                     /db_xref="taxon:10090"
                     /clone="C0037E03"
                     /clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA
                     library"
                     /sex="unknown"
                     /tissue_type="ectoplacental cone"
                     /dev_stage="embryonic day 7.5 postconception"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT1 (Life Technologies);
                     Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
                     ectoplacental cone of 7.5-dpc embryos.  The
                     double-stranded cDNA was synthesized from total RNAs with
                     an Oligo(dT) primer. The library was constructed by Minoru
                     S. H. Ko."
BASE COUNT       51 a      83 c      78 g      54 t       1 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 33;  Length 267;
  Best Local Similarity   88.9%;  Pred. No. 2.2e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| ||||| ||
Db      36 GGCGGCGAGTCTGCCGGC 19


RESULT   9
AA685482/c
LOCUS       AA685482      319 bp    mRNA            EST       03-APR-1998
DEFINITION  EST107297 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCBS54
            5' end similar to Ribosomal protein L8, mRNA sequence.
ACCESSION   AA685482
VERSION     AA685482.1  GI:2672080
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE     Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  MEDLINE   95396786
COMMENT     Other_ESTs: TC44630
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2002002"
                     /db_xref="taxon:10118"
                     /clone="RPCBS54"
                     /clone_lib="Rat PC-12 cells, untreated"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; poly(A)+ RNA was purified from untreated PC12 cells
                     cultured for 9 days.  cDNA was constructed using an
                     oligo-dT primer and directionally cloned using the Lambda
                     ZAP II Vector Kit by Stratagene"
BASE COUNT       67 a      86 c      97 g      66 t       3 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 37;  Length 319;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           |||||||||| ||||| ||
Db      25 GGCGGCGAGTCTGCCGGC 8


RESULT  10
AI131683/c
LOCUS       AI131683      333 bp    mRNA            EST       14-SEP-1998
DEFINITION  ue95e06.x1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:1498882 3' similar to gb:M64716 40S RIBOSOMAL PROTEIN S25
```

```
            (HUMAN);, mRNA sequence.
ACCESSION   AI131683
VERSION     AI131683.1  GI:3601699
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1402245.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:936486
            Trace considered overall poor quality
            Seq primer: custom primer used
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .333
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1498882"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT};
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT       78 a     78 c     75 g    102 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 42;  Length 333;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          | ||||||||||| |||||
Db    313 GCCGGCGAGTCGGCCGGC 296


RESULT  11
W09281/c
LOCUS       W09281         345 bp    mRNA            EST       01-OCT-1997
DEFINITION  ma06b04.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:303727 5' similar to gb:Z28407 60S RIBOSOMAL PROTEIN L8
            (HUMAN);, mRNA sequence.
ACCESSION   W09281
VERSION     W09281.1  GI:1283597
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692458.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 200487
            Seq primer: mob.REGA+ET
            High quality sequence stop: 201.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:303727"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by,
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       78 a     98 c     96 g     73 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 25;  Length 345;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||| |||||
Db     26 GGCGGCGAGTCTGCCGGC 9


RESULT  12
W91633/c
LOCUS       W91633         360 bp    mRNA            EST       09-JUL-1996
DEFINITION  MTA.H10.085.B MTA adult mouse thymus library Mus musculus cDNA
            clone MTA.H10.085 5' end similar to rat ribo prot L8, mRNA
            sequence.
ACCESSION   W91633
VERSION     W91633.1  GI:1407894
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Nguyen,C., Rocha,D., Granjeaud,S., Bernard,K., Naquet,P. and
```

```
            Jordan,B.R.
  TITLE     Gene expression in different cell types of the mouse thymus
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785693.
            Contact: Jordan BR
            Genome Structure and Immune Functions
            Centre d'Immunologie INSERM/CNRS
            Case 906, 13288 MARSEILLE Cedex 9, FRANCE
            Tel: 330491269496
            Fax: 330491269430
            Email: jordan@ciml.univ-mrs.fr
            This sequence was determined at HGMP Resource Centre, Hinxton,
            United Kingdom (M Botcherby)
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Mus musculus"
                     /strain="C57Bl/6"
                     /db_xref="taxon:10090"
                     /clone="MTA.H10.085"
                     /clone_lib="MTA adult mouse thymus library"
                     /lab_host="MC1061 p3"
                     /note="Vector: pcDNA1; Site_1: Not1; Site_2: EcoR1; The
                     cDNA library was constructed from poly(A)+ RNA of an adult
                     mouse  thymus by oligo-dT primed reverse transcription.
                     cDNA was selected on  gel for size above 800 nucleotides
                     after second strand synthesis, then  directionally cloned
                     after adaptor ligation and restriction enzyme  cleavage
                     into the pcDNA1 vector (Not 1 on polyA side, EcoR 1 on the
                     5'side)."
BASE COUNT       72 a    103 c    116 g     69 t
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 26;  Length 360;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db      21 GGCGGCGAGTCTGCCGGC 4


RESULT  13
AW229945/c
LOCUS       AW229945      360 bp    mRNA            EST       10-DEC-1999
DEFINITION  uo45b04.y1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:2645455 5'
            similar to gb:Z28407 60S RIBOSOMAL PROTEIN L8 (HUMAN);, mRNA
            sequence.
ACCESSION   AW229945
VERSION     AW229945.1  GI:6559241
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5420752.
            Other_ESTs: uo45b04.x2
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            MGI:1025907
            Seq primer: -40RP from Gibco
            High quality sequence stop: 301.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2645455"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       72 a    103 c    114 g     70 t      1 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 74;  Length 360;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db      18 GGCGGCGAGTCTGCCGGC 1


RESULT  14
AU032296
LOCUS       AU032296      363 bp    mRNA            EST       20-OCT-1998
DEFINITION  AU032296 Rice root Oryza sativa cDNA clone R3854_3A, mRNA sequence.
ACCESSION   AU032296
VERSION     AU032296.1  GI:3768269
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Minobe,Y. and Sasaki,T.
  TITLE     Rice cDNA from root
  JOURNAL   Unpublished (1995)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1798617.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R3854_3A"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT       42 a    142 c    121 g     56 t      2 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 43;  Length 363;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||| ||||| ||
Db     143 GGCGGCGAGGCGACCTGC 160


RESULT  15
AI625890
LOCUS       AI625890      380 bp    mRNA           EST       05-MAY-1999
DEFINITION  CrESTz0016 Chlamydomonas Uni-Zap XR Chlamydomonas reinhardtii cDNA
            3', mRNA sequence.
ACCESSION   AI625890
VERSION     AI625890.1  GI:4662690
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Purton,S. and Madagan,K.
  TITLE     Chlamydomonas reinhardtii ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188347.
            Contact: Saul Purton, Kathryn Madagan
            Department of Biology
            University College London
            Darwin Building, Gower Street, London, WC1E 6BT.
            Tel: 01713807098
            Fax: 01713807096
            Email: k.madagan@ucl.ac.uk, s.purton@ucl.ac.uk
            Similar to human ribosomal protein L19
            PCR PRimers
            FORWARD: TGGGAAGGGCGATCGGTGCGGGC
            BACKWARD: GAAAGCGGGCAGTGAGCGCAAGCG
            Seq primer: CGAAATTAACCCTCACTAAAGGG.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Chlamydomonas reinhardtii"
                     /strain="wild type 137C"
                     /db_xref="taxon:3055"
                     /clone_lib="Chlamydomonas Uni-Zap XR"
                     /sex="mt+"
                     /note="Vector: lambda Uni-Zap XR; Site_1: EcoRI; Site_2:
                     XhoI; The library was grown from cells that were adapting
                     to low CO2 conditions. They were growing auxotrophically
                     on elevated CO2, then dropped to air levels of CO2. The
                     cells were harvested 2 and 2.5 hours later."
BASE COUNT       84 a    109 c    122 g     64 t      1 others
ORIGIN


  Query Match             82.2%;  Score 14.8;  DB 48;  Length 380;
  Best Local Similarity   88.9%;  Pred. No. 2.3e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           || ||| ||||||||||
Db     219 GGTGGCAAGTCGACCGGC 236


Search completed: June 22, 2000, 17:43:45
Job time: 20095 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:43 ; Search time 783.2 Seconds
(without alignments)
-29.810 Million cell updates/sec

Title: US-09-362-485-23
Perfect score: 24
Sequence: 1 CGAGACTAATTTGGGTGCCTTGGC 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 17 5538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	100.0	24	5	A87626	A87626 Sequence 23
c	2	24	100.0	1194	2	MTU92472	U92472 Mycobacteri
c	3	24	100.0	1208	5	A89752	A89752 Sequence 9
c	4	24	100.0	1209	5	A87612	A87612 Sequence 9
c	5	24	100.0	1228	5	A87608	A87608 Sequence 5
c	6	24	100.0	1228	5	A89748	A89748 Sequence 5
c	7	24	100.0	1229	5	A87610	A87610 Sequence 7
c	8	24	100.0	1229	5	A89750	A89750 Sequence 7
c	9	24	100.0	1235	5	A87606	A87606 Sequence 3
c	10	24	100.0	1235	5	A87609	A87609 Sequence 6
c	11	24	100.0	1235	5	A87611	A87611 Sequence 8
c	12	24	100.0	1235	5	A89746	A89746 Sequence 3
c	13	24	100.0	1235	5	A89749	A89749 Sequence 6
c	14	24	100.0	1235	5	A89751	A89751 Sequence 8
c	15	24	100.0	1236	5	A87613	A87613 Sequence 10
c	16	24	100.0	1236	5	A89753	A89753 Sequence 10
c	17	24	100.0	1237	5	A87607	A87607 Sequence 4
c	18	24	100.0	1237	5	A89747	A89747 Sequence 4
c	19	24	100.0	1245	5	A87605	A87605 Sequence 2
c	20	24	100.0	1245	5	A89745	A89745 Sequence 2
c	21	24	100.0	1260	5	A87604	A87604 Sequence 1
c	22	24	100.0	1260	5	A89744	A89744 Sequence 1
c	23	24	100.0	2584	1	MTALADH	X63069 M.tuberculo
c	24	24	100.0	56414	1	MTV002	AL008967 Mycobacte
	25	18.2	75.8	2349	8	AF046923	AF046923 Colletotr
c	26	18.2	75.8	78947	32	AP000575	AP000575 Homo sapi
c	27	18.2	75.8	109290	32	HS838L14	Y12335 Homo sapien
c	28	18.2	75.8	115381	41	AC009812	AC009812 Homo sapi
	29	18.2	75.8	148996	52	AC016696	AC016696 Homo sapi
c	30	18.2	75.8	180898	32	AP000928	AP000928 Homo sapi
	31	18.2	75.8	188211	45	AC009871	AC009871 Homo sapi
	32	18.2	75.8	203300	40	AC000134	AC000134 Homo sapi
	33	17.6	73.3	444	10	HSPROL5	X00542 H.sapiens g
	34	17.6	73.3	83538	43	AC018338	AC018338 Drosophil
	35	17.6	73.3	105214	43	AC008183	AC008183 Drosophil
	36	17.6	73.3	114862	42	AC014966	AC014966 Drosophil
	37	17.6	73.3	118145	56	AC010664	AC010664 Drosophil
c	38	17.6	73.3	177791	52	AC016781	AC016781 Oryza sat
	39	17.4	72.5	78988	43	AC016790	AC016790 Homo sapi
	40	17.4	72.5	225958	33	AL138797	AL138797 Homo sapi
c	41	17.2	71.7	1395	2	AF043282	AF043282 Streptoco
	42	17.2	71.7	2573	7	YSCDPB2	M61710 S.cerevisia
	43	17.2	71.7	3024	11	AF009014	AF009014 Homo sapi
	44	17.2	71.7	3361	7	SCYGR198W	Z72983 S.cerevisia
	45	17.2	71.7	9000	7	SCG7914	X82775 S.cerevisia

ALIGNMENTS

```
RESULT   1
A87626
LOCUS       A87626          24 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 23 from Patent WO9836089.
ACCESSION   A87626
VERSION     A87626.1  GI:6736266
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      5 c      8 g      7 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db      1 CGAGACTAATTTGGGTGCCTTGGC 24


RESULT   2
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 2;  Length 1194;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT   3
A89752/c
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT   4
A87612/c
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT   5
A87608/c
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT   6
A89748/c
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT   7
A87610/c
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCTTGGC 789


RESULT   8
A89750/c
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCTTGGC 789


RESULT   9
A87606/c
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
```

```
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT  10
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCTTGGC 789


RESULT  11
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCT GGC 789


RESULT  12
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790


RESULT  13
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

```
  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCTTGGC 789


RESULT  14
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    812 CGAGACTAATTTGGGTGCCTTGGC 789


RESULT  15
A87613/c
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 24;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGAGACTAATTTGGGTGCCTTGGC 24
          ||||||||||||||||||||||||
Db    813 CGAGACTAATTTGGGTGCCTTGGC 790
```

Search completed: June 22, 2000, 17:58:45
Job time: 10794 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:01:53 ; Search time 175.2 Seconds
(without alignments)
34.273 Million cell updates/sec

Title: US-09-362-485-23
Perfect score: 24
Sequence: 1 CGAGACTAATTTGGGTGCCTTGGC 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		24	100.0	24	1	V49524	Mycobacterium sp.
2		24	100.0	24	1	V49622	AlaDH derived olig
3	c	24	100.0	1245	1	V49626	Mycobacterium tube
4	c	24	100.0	1260	1	V49510	Mycobacterium sp.
5	c	24	100.0	1260	1	V49625	Mycobacterium tube
6	c	16.8	70.0	[illegible]7	1	T27948	Hepatitis C virus
7	c	16.6	69.2	191	1	T23025	Human gene signatu
8		16.2	67.5	5579	1	T30056	S. pneumoniae dete
9	c	16.2	67.5	14872	1	V52205	Streptococcus pneu
10		16	66.7	16	1	V49525	Mycobacterium sp.
11		16	66.7	16	1	V49623	AlaDH derived olig
12		15.8	65.8	622	1	X30872	Streptococcus pneu
13		15.8	65.8	653	1	V52491	Streptococcus pneu
14	c	15.8	65.8	2061	1	V57439	Human protein kina
15	c	15.8	65.8	2327	1	V57440	Human protein kina
16	c	15.8	65.8	7011	1	V20464	Human L-myc oncoge
17		15.8	65.8	12566	1	V52282	Streptococcus pneu
18	c	15.6	65.0	101	1	V02808	Human RHAMM genomi
19	c	15.6	65.0	3114	1	V02800	Human RHAMM cDNA.
20	c	15.6	65.0	3248	1	V27580	Human hepatocyte n
21	c	15.6	65.0	4325	1	Q81136	RPLA2-8 gene. Nove
22	c	15.2	63.3	3831	1	Q11873	appCA region encod
23	c	15.2	63.3	4731	1	T73219	ApxIIAB'C gene. Pr
24		15.2	63.3	8834	1	Q13874	RgR gene. DNA enco
25		15.2	63.3	19932	1	T46159	CagI locus. Helico
26	c	15	62.5	555	1	X13583	Enterococcus faeca
27		15	62.5	786	1	Q51273	Acidic protease is
28	c	15	62.5	1089	1	N91027	DNA sequence in pl
29	c	15	62.5	1143	1	N70103	Sequence encoding
30		15	62.5	1453	1	Q51277	Acidic protease V2
31	c	15	62.5	1866	1	Q04292	Encodes polypeptid
32		15	62.5	2225	1	T85144	Human Immune Cell
33		15	62.5	18627	1	V52246	Streptococcus pneu
34		15	62.5	40352	1	V02032	MAGE-B cluster DNA
35	c	15	62.5	50341	1	V22674	DNA sequence of a
36	c	15	62.5	52297	1	T51411	Mycobacteriophage
37	c	15	62.5	52298	1	Q47357	L5 mycobacteriopha
38	c	14.8	61.7	1260	1	V62753	Human secreted pro
39	c	14.8	61.7	1269	1	X22121	Human secreted pro
40	c	14.8	61.7	3438	1	Q05769	GALP3 Promoter. Re
41		14.6	60.8	108	1	T58937	PSLGL-mature LGH t
42	c	14.6	60.8	110	1	Q04068	Endothelin-2 encod
43	c	14.6	60.8	110	1	Q42528	Synthetic DNA enco
44	c	14.6	60.8	229	1	T20128	Human gene signatu
45		14.6	60.8	573	1	V53348	DNA encoding a Sta

ALIGNMENTS

RESULT 1
V49524
ID V49524 standard; DNA; 24 BP.
AC V49524;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R6.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic ac d encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 24 BP; 4 A; 5 C; 8 G; 7 T;

Query Match 100.0%; Score 24; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGAGACTAATTTGGGTGCCTTGGC 24
           ||||||||||||||||||||||||
Db       1 CGAGACTAATTTGGGTGCCTTGGC 24
```

RESULT 2
V49622
ID V49622 standard; DNA; 24 BP.
AC V49622;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R6.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 24 BP; 4 A; 5 C; 8 G; 7 T;

Query Match 100.0%; Score 24; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
||||||||||||||||||||||||
Db 1 CGAGACTAATTTGGGTGCCTTGGC 24

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 24; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
||||||||||||||||||||||||
Db 819 CGAGACTAATTTGGGTGCCTTGGC 796

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 24; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
||||||||||||||||||||||||
Db 834 CGAGACTAATTTGGGTGCCTTGGC 811

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 24; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
||||||||||||||||||||||||
Db 834 CGAGACTAATTTGGGTGCCTTGGC 811

RESULT 6
T27948/c
ID T27948 standard; DNA; 447 BP.
AC T27948;
DT 10-MAR-1997 (first entry)
DE Hepatitis C virus type 2h isolate BNL5 bases 478-925.
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
KW PCR; primer; probe; antibody; infection; ss.
OS Hepatitis C virus.
PN WO9613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI; 96-251460/25.
DR P-PSDB; R96537.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 6; Fig 3; 150pp; English.
CC The sequences T27937-T27989 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 21,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
CC This sequence represents nucleotides 478-975 from the HCV type 2h isolate
CC BNL5.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 447 BP; 86 A; 129 C; 111 G; 121 T;

Query Match 70.0%; Score 16.8; DB _; Length 447;
Best Local Similarity 90.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCT 20
||||||| |||| |||||||
Db 233 CGAGACTGATTTTGGTGCCT 214

RESULT 7
T23025/c
ID T23025 standard; cDNA to mRNA; 191 BP.
AC T23025;
DT 23-AUG-1996 (first entry)
DE Human gene signature HUMGS04754.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1272; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 191 BP; 42 A; 53 C; 51 G; 44 T;

Query Match 69.2%; Score 16.6; DB 1; Length 191;
Best Local Similarity 79.2%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
| ||||| ||||||| ||||||
Db 45 CCAGACTCCCTTGGGTGNCTTGGC 22

RESULT 8
T30056
ID T30056 standard; DNA; 5579 BP.
AC T30056;
DT 21-NOV-1996 (first entry)
DE S. pneumoniae detection probe, SP-6-6.
KW PstI fragment; S. pneumoniae; probe; detection; infectious disease;
KW pneumonia; NYSDH DP-2; S. agalactiae IFM 58/59; S. anginosus;
KW Staphylococcus aureus ATCC 25923; E. coli ATCC 25922;
KW Candida albicans IFM 4008A; Klebsiella pneumoniae; diagnosis; ds.
OS Streptococcus pneumoniae.
PN WO9610647-A1.
PD 11-APR-1996.
PF 02-OCT-1995; J02036.
PR 30-SEP-1994; JP-236348.
PA (FUSO) FUSO PHARM IND LTD.
PI Eda S, Keshi H, Matsuhisa A, Uehara H, Ueyama H;
DR WPI; 96-209367/21.
PT Probes for the diagnosis of pneumonia caused by Streptococcus
PT pneumoniae - isolated from PstI cleavage of the S. pneumoniae
PT genome.
PS Claim 2; Page 27-31; 39pp; Japanese.
CC The sequences given in T30050-56 are PstI fragments of the S.
CC pneumoniae genome which may be used as probes in the detection of
CC infectious diseases, esp. pneumonia. These probes were isolated

CC from S. pneumoniae cells which were cultured on heart-brain infusion
CC medium. These probes hybridised with genomic DNA from S. pneumoniae
CC NYSDH DP-2 and two clinical isolates of S. pneumoniae. They did not
CC react with DNA isolated from S. agalactiae IFM 58/59, S. anginosus
CC or nine other strains of Streptococcus, Staphylococcus aureus ATCC
CC 25923, E. coli ATCC 25922, Candida albicans IFM 4008A, Klebsiella
CC pneumoniae or human genomic DNA. The probes provide accurate and
CC specific diagnosis of S. pneumoniae pneumonia by hybridisation of the
CC probe with genomic DNA contained in a clinical sample.
SQ Sequence 5579 BP; 1716 A; 1219 C; 1015 G; 1629 T;

Query Match 67.5%; Score 16.2; DB 1; Length 5579;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGACTAATTTGGGTGCCTTG 22
|||| ||||||||| |||||
Db 5167 GAGAATAATTTGGGCTCCTTG 5187

RESULT 9
V52205/c
ID V52205 standard; DNA; 14872 BP.
AC V52205;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:72.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 593-602; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 14872 BP; 4340 A; 2874 C; 3210 G; 4448 T;

Query Match 67.5%; Score 16.2; DB 1; Length 14872;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGACTAATTTGGGTGCCTTG 22
|||| ||||||||| |||||
Db 12122 GAGAATAATTTGGGCTCCTTG 12102

RESULT 10
V49525
ID V49525 standard; DNA; 16 BP.
AC V49525;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R7.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 16 BP; 1 A; 3 C; 6 G; 6 T;

Query Match 66.7%; Score 16; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ATTTGGGTGCCTTGGC 24
||||||||||||||||
Db 1 ATTTGGGTGCCTTGGC 16

RESULT 11
V49623
ID V49623 standard; DNA; 16 BP.
AC V49623;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R7.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to

```
CC   identify substances that inhibit mycobacteria, for combatting epidemics
CC   and for vaccination follow-up. The oligonucleotides are used similarly
CC   in diagnostic hybridisation tests, also for culture confirmation of
CC   isolated strains and for chromosome fingerprinting to detect/
CC   differentiate between mycobacteria, and for L-alanine-specific
CC   biotransformation reactions. AlaDH is an early antigen, present
CC   extracellularly after only a few days of growth, making it an ideal
CC   drug target.
SQ   Sequence   16 BP;    1 A;    3 C;    6 G;    6 T;


  Query Match            66.7%;  Score 16;  DB 1;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 23;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 ATTTGGGTGCCTTGGC 24
          ||||||||||||||||
Db      1 ATTTGGGTGCCTTGGC 16


RESULT  12
X30872
ID   X30872 standard; DNA; 622 BP.
AC   X30872;
DT   20-MAY-1999  (first entry)
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:149.
KW   Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
OS   Streptococcus pneumoniae.
PN   WO9737026-A1.
PD   09-OCT-1997.
PF   01-APR-1997; U05306.
PR   22-AUG-1996; US-025788.
PR   02-APR-1996; US-014690.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Stodola RK;
DR   WPI; 97-503111/46.
DR   P-PSDB; Y11289.
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
PS   Claim 5; Page 146; 354pp; English.
CC   X30724 to X30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC   encode the novel proteins given in Y11114 to Y11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
CC   be used to treat an individual having need to inhibit a bacterial
CC   protein. Vectors expressing the proteins can be used to induce a
CC   protective immune response in mammals.
SQ   Sequence   622 BP;    162 A;    115 C;    166 G;    177 T;


  Query Match            65.8%;  Score 15.8;  DB 1;  Length 622;
  Best Local Similarity  89.5%;  Pred. No. 47;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 ACTAATTTGGGTGCCTTGG 23
          |||| ||| ||||||||||
Db    454 ACTATTTTCGGTGCCTTGG 472


RESULT  13
V52491
ID   V52491 standard; DNA; 653 BP.
AC   V52491;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:358.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI; 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 1372; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence   653 BP;    192 A;    127 C;    154 G;    180 T;


  Query Match            65.8%;  Score 15.8;  DB 1;  Length 653;
  Best Local Similarity  89.5%;  Pred. No. 47;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 ACTAATTTGGGTGCCTTGG 23
          |||| ||| ||||||||||
Db     97 ACTATTTTCGGTGCCTTGG 115


RESULT  14
V57439/c
ID   V57439 standard; cDNA; 2061 BP.
AC   V57439;
DT   21-DEC-1998  (first entry)
DE   Human protein kinase hYAK3-alpha cDNA.
KW   Protein kinase; hYAK-alpha; human; bone loss; osteoporosis;
KW   inflammation; acute respiratory distress syndrome; osteoarthritis;
KW   rheumatoid arthritis; inflammatory bowel disease; psoriasis;
KW   dermatitis; asthma; allergy; infection; HIV; immunodeficiency;
KW   cachexia; septic shock; pain; injury; cancer; testicular cancer;
KW   anorexia, bulimia; Parkinson's disease; cardiovascular disease;
KW   restenosis; atherosclerosis; heart failure; myocardial infarction;
KW   hypotension; hypertension; urinary retention; angina pectoris;
KW   ulcer; benign prostatic hypertrophy; psychosis; schizophrenia;
KW   anxiety; manic depression; delirium; severe mental retardation;
KW   dementia; Huntington's disease; Gilles de la Tourette syndrome;
KW   dyskinesia; diagnosis; therapy; vaccine; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             146. .1912
FT                   /*tag= a
PN   EP-870825-A1.
PD   14-OCT-1998.
PF   05-MAR-1998; 301641.
PR   07-APR-1997; US-835170.
```

PR 05-MAR-1997; US-040618.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Creasy CL, Xie W;
DR WPI; 98-523155/45.
DR P-PSDB; W75790.
PT New DNA encoding hYAK3 human protein kinase polypeptides - used to
PT treat and diagnose e.g. osteoporosis, inflammatory diseases,
PT inflammatory bowel disease, psoriasis, dermatitis, asthma,
PT allergies, infections, septic shock, anorexia and cancer
PS Claim 5; Fig 1; 23pp; English.
CC This full-length cDNA clone codes for a novel human protein kinase,
CC termed hYAK3-alpha (see W75790), which is expressed predominantly
CC in skeletal muscle. A partial clone was initially identified in a
CC search of the Human Genome Sciences database on the basis of
CC homology to yeast YAK1. This partial clone was used to screen
CC human testis and skeletal muscle cDNA libraries. hYAK3-alpha cDNA
CC was obtained from the skeletal muscle library, and hYAK3-beta cDNA
CC (see V57440) from the testis library. hYAK-alpha cDNA is 266
CC nucleotides shorter that hYAK-beta cDNA and the 3' most 1844
CC nucleotides are identical. The differences between the 2 cDNAs may
CC be due to alternative splicing. The invention provides host cells
CC and methods for producing hYAK3 polypeptides, as well as methods
CC for treating subjects having need to enhance or inhibit hYAK3
CC activity, for diagnosing a disease related to expression or
CC activity of hYAK3, methods for identifying agonist and antagonist
CC compounds, and polynucleotide probes and primers. hYAK3
CC polypeptides and polynucleotides can be used to treat and diagnose
CC bone loss (e.g. osteoporosis), inflammatory diseases (e.g. ARDS),
CC rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,
CC psoriasis, dermatitis, asthma, allergies, infections (e.g.
CC bacterial, fungal, protozoal and viral infections such as HIV),
CC HIV-associated cachexia and other immunodeficiency disorders,
CC septic shock, pain, injury, cancers (e.g. testicular cancer),
CC anorexia, bulimia, Parkinson's disease, cardiovascular disease
CC (e.g. restenosis, atherosclerosis, acute heart failure and
CC myocardial infarction), hypotension, hypertension, urinary
CC retention, angina pectoris, ulcers, benign prostatic hypertrophy
CC and psychotic and neurological disorders (e.g. anxiety,
CC schizophrenia, manic depression, delirium, dementia, severe mental
CC retardation and dyskinesias such as Huntington's disease or Gilles
CC de la Tourette syndrome).
SQ Sequence 2061 BP; 574 A; 471 C; 514 G; 502 T;

Query Match 65.8%; Score 15.8; DB 1; Length 2061;
Best Local Similarity 89.5%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ACTAATTTGGGTGCCTTGG 23
|| | ||||||||||||||
Db 561 ACCACTTTGGGTGCCTTGG 543

RESULT 15
V57440/c
ID V57440 standard; cDNA; 2327 BP.
AC V57440;
DT 21-DEC-1998 (first entry)
DE Human protein kinase hYAK3-beta cDNA.
KW Protein kinase; hYAK-beta; human; bone loss; osteoporosis;
KW inflammation; acute respiratory distress syndrome; osteoarthritis;
KW rheumatoid arthritis; inflammatory bowel disease; psoriasis;
KW dermatitis; asthma; allergy; infection; HIV; immunodeficiency;
KW cachexia; septic shock; pain; injury; cancer; testicular cancer;
KW anorexia, bulimia; Parkinson's disease; cardiovascular disease;
KW restenosis; atherosclerosis; heart failure; myocardial infarction;
KW hypotension; hypertension; urinary retention; angina pectoris;
KW ulcer; benign prostatic hypertrophy; psychosis; schizophrenia;
KW anxiety; manic depression; delirium; severe mental retardation;
KW dementia; Huntington's disease; Gilles de la Tourette syndrome;
KW dyskinesia; diagnosis; therapy; vaccine; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 472. .2178
FT /*tag= a
PN EP-870825-A1.
PD 14-OCT-1998.
PF 05-MAR-1998; 301641.
PR 07-APR-1997; US-835170.
PR 05-MAR-1997; US-040618.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Creasy CL, Xie W;
DR WPI; 98-523155/45.
DR P-PSDB; W75791.
PT New DNA encoding hYAK3 human protein kinase polypeptides - used to
PT treat and diagnose e.g. osteoporosis, inflammatory diseases,
PT inflammatory bowel disease, psoriasis, dermatitis, asthma,
PT allergies, infections, septic shock, anorexia and cancer
PS Claim 5; Fig 2; 23pp; English.
CC This full-length cDNA clone codes for a novel human protein kinase,
CC termed hYAK3-beta (see W757901, which is expressed predominantly
CC in testis. A partial clone was initially identified in a search of
CC the Human Genome Sciences database on the basis of homology to
CC yeast YAK1. This partial clone was used to screen human testis and
CC skeletal muscle cDNA libraries. hYAK3-beta cDNA was obtained from
CC the testis cDNA library, and hYAK3-alpha cDNA (see V57439) from the
CC skeletal muscle library. hYAK-beta cDNA is 266 nucleotides longer
CC than hYAK-alpha cDNA and the 3' most 1844 nucleotides are
CC identical. The differences between the 2 cDNAs may be due to
CC alternative splicing. The invention provides host cells and
CC methods for producing hYAK3 polypeptides, as well as methods for
CC treating subjects having need to enhance or inhibit hYAK3 activity,
CC for diagnosing a disease related to expression or activity of
CC hYAK3, methods for identifying agonist and antagonist compounds,
CC and polynucleotide probes and primers. hYAK3-alpha and -beta
CC polypeptides and polynucleotides can be used to treat and diagnose
CC bone loss (e.g. osteoporosis), inflammatory diseases (e.g. ARDS),
CC rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,
CC psoriasis, dermatitis, asthma, allergies, infections (e.g.
CC bacterial, fungal, protozoal and viral infections such as HIV),
CC HIV-associated cachexia and other immunodeficiency disorders,
CC septic shock, pain, injury, cancers (e.g. testicular cancer),
CC anorexia, bulimia, Parkinson's disease, cardiovascular disease
CC (e.g. restenosis, atherosclerosis, acute heart failure and
CC myocardial infarction), hypotension, hypertension, urinary
CC retention, angina pectoris, ulcers, benign prostatic hypertrophy
CC and psychotic and neurological disorders (e.g. anxiety,
CC schizophrenia, manic depression, delirium, dementia, severe mental
CC retardation and dyskinesias such as Huntington's disease or Gilles
CC de la Tourette syndrome).
SQ Sequence 2327 BP; 660 A; 515 C; 573 G; 579 T;

Query Match 65.8%; Score 15.8; DB 1; Length 2327;
Best Local Similarity 89.5%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ACTAATTTGGGTGCCTTGG 23
|| | ||||||||||||||
Db 827 ACCACTTTGGGTGCCTTGG 809

Search completed: June 22, 2000, 18:01:55
Job time: 10127 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:04:01 ; Search time 123.79 Seconds
(without alignments)
25.201 Million cell updates/sec

Title: US-09-362-485-23
Perfect score: 24
Sequence: 1 CGAGACTAATTTGGGTGCCTTGGC 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	16.2	67.5	363	3	US-08-967-101-30	Sequence 30, Appl
c	2	16.2	67.5	363	4	US-08-592-541-30	Sequence 30, Appl
c	3	15.8	65.8	2061	4	US-08-835-170-1	Sequence 1, Appli
c	4	15.8	65.8	2327	4	US-08-835-170-3	Sequence 3, Appli
c	5	15.8	65.8	7011	2	US-08-306-691B-42	Sequence 42, Appl
c	6	15.6	65.0	4325	4	US-08-888-497-21	Sequence 21, Appl
c	7	15.6	65.0	4325	6	PCT-US94-07926-21	Sequence 21, Appl
c	8	15.2	63.3	4731	4	US-08-488-706-2	Sequence 2, Appli
c	9	15.2	63.3	4731	5	US-08-772-270A-9	Sequence 9, Appli
	10	15.2	63.3	19932	4	US-08-477-451-25	Sequence 25, Appl
	11	15	62.5	2225	4	US-08-780-370A-1	Sequence 1, Appli
	12	15	62.5	40352	5	US-08-846-111D-15	Sequence 15, Appl
c	13	15	62.5	50341	2	US-08-247-901C-1	Sequence 1, Appli
c	14	15	62.5	50341	4	US-09-075-904-1	Sequence 1, Appli
c	15	14.6	60.8	99	2	US-08-673-269-9	Sequence 9, Appli
c	16	14.6	60.8	110	7	5231166-2	Patent No. 5231166
	17	14.6	60.8	1104	1	US-08-423-399B-36	Sequence 36, Appl
	18	14.6	60.8	1162	1	US-08-423-399B-34	Sequence 34, Appl
c	19	14.6	60.8	1218	2	US-08-673-269-1	Sequence 1, Appli
	20	14.6	60.8	1455	1	US-08-423-399B-32	Sequence 32, Appl
	21	14.6	60.8	1455	2	US-08-446-803-5	Sequence 5, Appli
	22	14.6	60.8	1455	3	US-08-861-837-5	Sequence 5, Appli
c	23	14.6	60.8	2397	4	US-08-978-404B-2	Sequence 2, Appli
	24	14.6	60.8	2407	3	US-08-895-522-2	Sequence 2, Appli
c	25	14.6	60.8	7620	1	US-07-767-135-1	Sequence 1, Appli
c	26	14.6	60.8	7620	1	US-07-841-652-1	Sequence 1, Appli
	27	14.4	60.0	108	1	US-08-468-674B-52	Sequence 52, Appl
	28	14.4	60.0	108	2	US-08-780-571-52	Sequence 52, Appl
c	29	14.4	60.0	1167	1	US-07-960-985-1	Sequence 1, Appli
c	30	14.4	60.0	1167	3	US-08-496-671-1	Sequence 1, Appli
	31	14.4	60.0	1230	5	US-08-793-666-14	Sequence 14, Appl
	32	14.4	60.0	2606	3	US-08-568-459A-7	Sequence 7, Appli
	33	14.4	60.0	2606	4	US-08-487-826B-7	Sequence 7, Appli
	34	14.4	60.0	49272	2	US-08-614-770A-1	Sequence 1, Appli
c	35	14.2	59.2	2676	4	US-08-511-485-5	Sequence 5, Appli
	36	14	58.3	299	4	US-08-606-639A-6	Sequence 6, Appli
	37	14	58.3	348	1	US-08-468-674B-72	Sequence 72, Appl
	38	14	58.3	348	2	US-08-780-571-72	Sequence 72, Appl
	39	14	58.3	394	3	US-07-885-089B-15	Sequence 15, Appl
	40	14	58.3	394	3	US-07-885-089B-17	Sequence 17, Appl
	41	14	58.3	459	3	US-08-852-807-19	Sequence 19, Appl
	42	14	58.3	615	3	US-08-748-640-1	Sequence 1, Appli
	43	14	58.3	615	3	US-08-686-349-1	Sequence 1, Appli
c	44	14	58.3	707	2	US-08-592-126-119	Sequence 119, App
c	45	14	58.3	1627	4	US-08-389-564B-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT   1
US-08-967-101-30/c
; Sequence 30, Application US/08967101
; Patent No. 5840540
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,101
;      FILING DATE:  10-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/592,541
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-967-101-30

  Query Match             67.5%;  Score 16.2;  DB 3;  Length 363;
  Best Local Similarity   85.7%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GACTAATTTGGGTGCCTTGGC 24
```

```
          ||||||||| |||||  ||||
Db      34 GACTAATTTTGGTGCAGTGGC 14


RESULT   2
US-08-592-541-30/c
; Sequence 30, Application US/08592541
; Patent No. 5986054
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,541
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-592-541-30


  Query Match           67.5%;  Score 16.2;  DB 4;  Length 363;
  Best Local Similarity  85.7%;  Pred. No. 12;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 GACTAATTTGGGTGCCTTGGC 24
          ||||||||| |||||  ||||
Db     34 GACTAATTTTGGTGCAGTGGC 14


RESULT   3
US-08-835-170-1/c
; Sequence 1, Application US/08835170
; Patent No. 5965420
;  GENERAL INFORMATION:
;    APPLICANT:  Creasy, Caretha
;    APPLICANT:  Wei Xie
;    TITLE OF INVENTION:  Human Protein Kinases hYAK3
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,170
;      FILING DATE:  26-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  GH50004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2061 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-835-170-1


  Query Match           65.8%;  Score 15.8;  DB 4;  Length 2061;
  Best Local Similarity  89.5%;  Pred. No. 25;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 ACTAATTTGGGTGCCTTGG 23
          || | |||||||||||||||
Db    561 ACCAC TTGGGTGCCTTGG 543


RESULT   4
US-08-835-170-3/c
; Sequence 3, Application US/08835170
; Patent No. 5965420
;  GENERAL INFORMATION:
;    APPLICANT:  Creasy, Caretha
;    APPLICANT:  Wei Xie
;    TITLE OF INVENTION:  Human Protein Kinases hYAK3
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,170
;      FILING DATE:  26-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  GH50004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
```

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other
US-08-835-170-3


  Query Match             65.8%;  Score 15.8;  DB 4;  Length 2327;
  Best Local Similarity   89.5%;  Pred. No. 25;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 ACTAATTTGGGTGCCTTGG 23
           || | ||||||||||||||||
Db     827 ACCACTTTGGGTGCCTTGG 809


RESULT   5
US-08-306-691B-42/c
; Sequence 42, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;    APPLICANT:  Calabretta, Bruno
;    APPLICANT:  Skorski, Tomasz
;    TITLE OF INVENTION:  ANTISENSE
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;      STREET:  Two Penn Center, Suite 1800
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,691B
;      FILING DATE:  September 15, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;      TELEX:  No. 5734039e
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-306-691B-42


  Query Match             65.8%;  Score 15.8;  DB 2;  Length 7011;
  Best Local Similarity   89.5%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 ACTAATTTGGGTGCCTTGG 23
           |||| |||||| |||||||
Db    4737 ACTACTTTGGGGGCCTTGG 4719


RESULT   6
US-08-888-497-21/c
; Sequence 21, Application US/08888497
; Patent No. 5972677
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,405
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  722..1195
US-08-888-497-21


  Query Match             65.0%;  Score 15.6;  DB 4;  Length 4325;
  Best Local Similarity   81.8%;  Pred. No. 36;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGG 23
             | ||  | ||||||||||||||||
Db    1938 GTGAGGACTTTGGGTGCCTTGG 1917


RESULT   7
PCT-US94-07926-21/c
; Sequence 21, Application PC/TUS9407926
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
```

```
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07926
;      FILING DATE:  15-JUL-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  722..1195
PCT-US94-07926-21


  Query Match           85.0%;  Score 15.6;  DB 6;  Length 4325;
  Best Local Similarity  81.8%;  Pred. No. 36;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGG 23
           | ||  | ||||||||||||||
Db    1938 GTGAGGACTTTGGGTGCCTTGG 1917


RESULT   8
US-08-488-706-2/c
; Sequence 2, Application US/08488706
; Patent No. 5994525
;  GENERAL INFORMATION:
;    APPLICANT:  Kamp, Elbarte M.
;    APPLICANT:  Smits, Marinus A.
;    TITLE OF INVENTION:  Recombinant Vaccine For Prevention and/or
;    TITLE OF INVENTION:  Treatment of Pleuropneumonia Infections
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abelman, Frayne & Schwab
;      STREET:  708 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch; 1.44 MB storage
;      COMPUTER:  IBM PC or compatible
;      OPERATING SYSTEM:  DOS 3.31
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,706
;      FILING DATE:  09-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/722,971
;      FILING DATE:  28-JUNE-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cinamon, Jay S.
;      REGISTRATION NUMBER:  24,156
;      REFERENCE/DOCKET NUMBER:  201,875
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-949-9022
;      TELEFAX:  212-949-9190
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4731
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-488-706-2


  Query Match           63.3%;  Score 15.2;  DB 4;  Length 4731;
  Best Local Similarity  85.0%;  Pred. No. 58;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GACTAATTTGGGTGCCTTGG 23
           ||| ||||| ||||| ||||
Db     309 GACCAATTTCGGTGCATTGG 290


RESULT   9
US-08-772-270A-9/c
; Sequence 9, Application US/08772270A
; Patent No. 6019984
;  GENERAL INFORMATION:
;    APPLICANT:  MacInnes, Janet
;    APPLICANT:  Ricciatti, Paul
;    APPLICANT:  Mallard, Bonnie
;    APPLICANT:  Rosendal, Soren
;    TITLE OF INVENTION:  NOVEL BACTERIAL PREPARATIONS, METHOD FOR
;    TITLE OF INVENTION:  PRODUCING SAME, AND THEIR USE AS VACCINES
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bereskin & Parr
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/772,270A
;      FILING DATE:  December 23, 1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  6580-81
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 364-7311
;      TELEFAX:  (416) 361-1398
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4731 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Actinobacillus pleuropneumoniae
US-08-772-270A-9


  Query Match             63.3%;  Score 15.2;  DB 5;  Length 4731;
  Best Local Similarity   85.0%;  Pred. No. 58;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GACTAATTTGGGTGCCTTGG 23
           ||| ||||| ||||| ||||
Db     309 GACCAATTTCGGTGCATTGG 290


RESULT  10
US-08-477-451-25
; Sequence 25, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19932 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-25


  Query Match             63.3%;  Score 15.2;  DB 4;  Length 19932;
  Best Local Similarity   85.0%;  Pred. No. 73;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 ACTAATTTGGGTGCCTTGGC 24
           || |||||| | ||||||||
Db    8113 ACCAATTTGCGCGCCTTGGC 8132


RESULT  11
US-08-780-370A-1
; Sequence 1, Application US/08780370A
; Patent No. 5962268
;  GENERAL INFORMATION:
;    APPLICANT:  SOPPET, DANIEL R
;    APPLICANT:  LI, YI
;    TITLE OF INVENTION:  IMMUNE CELL CYTOKINE
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GEN ME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/780,370A
;      FILING DATE:  09-JAN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMAT N:
;      NAME:  BROOKES, ALLAN A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF229
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  368..1330
US-08-780-370A-1


  Query Match             62.5%;  Score 15;  DB 4;  Length 2225;
  Best Local Similarity   78.3%;  Pred. No. 64;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGGC 24
           |||||| |  | ||||||||||
Db    1363 GAGACTTCTGGGTGTGCCTTGGC 1385


RESULT  12
US-08-846-111D-15
; Sequence 15, Application US/08846111D
; Patent No. 6017705
;  GENERAL INFORMATION:
;    APPLICANT:  Lurquin, Christophe; Brasseur, Francis;
;    APPLICANT:  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Which Are
;    TITLE OF INVENTION:  Members Of The Mage-B Family and Uses Thereof
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect For DOS 6.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,111D
;      FILING DATE:  25-APRIL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/658,578
```

```
;       FILING DATE:  5-MAY-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/403,388
;       FILING DATE:  14-MARCH-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, No. 6017705man D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 5444.1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  40352 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-846-111D-15


  Query Match             62.5%;  Score 15;  DB 5;  Length 40352;
  Best Local Similarity   78.3%;  Pred. No. 1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGG 23
           | || ||||| ||||||| || |
Db   14005 CTAGGCTAATGTGGGTGCTTTTG 14027


RESULT  13
US-08-247-901C-1/c
; Sequence 1, Application US/08247901C
; Patent No. 5750384
;  GENERAL INFORMATION:
;     APPLICANT:  Jacobs et al
;     TITLE OF INVENTION:  L5 SHUTTLE PHASMIDS
;     NUMBER OF SEQUENCES:  1
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Amster, Rothstein & Ebenstein
;       STREET:  90 Park Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10016
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Word Processor (ASCII)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/247,901C
;       FILING DATE:  May 23, 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/057,531
;       FILING DATE:  April 29, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bogosian, Elizabeth A
;       REGISTRATION NUMBER:  39,911
;       REFERENCE/DOCKET NUMBER:  96700/273
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 697-5995
;       TELEFAX:  (212) 286-0854 or 286-0082
;       TELEX:  TWX 710-581-4766
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  50341
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  L5 shuttle phasmid sequence
;     HYPOTHETICAL:  No
;     ANTI-SENSE:
;     FRAGMENT TYPE:
;     ORIGINAL SOURCE:
;       ORGANISM:  L5 mycobacteriophage
;       STRAIN:
;       INDIVIDUAL ISOLATE:
;       DEVELOPMENTAL STAGE:
;       HAPLOTYPE:
;       TISSUE TYPE:
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:  No. 5750384e
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1


  Query Match             62.5%;  Score 15;  DB 2;  Length 50341;
  Best Local Similarity   78.3%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGG 23
           || |  | | ||||||||||||
Db   15790 CGGGTATGACTTGGGTGCCTTGG 15768


RESULT  14
US-09-075-904-1/c
; Sequence 1, Application US/09075904
; Patent No. 5994137
;  GENERAL INFORMATION:
;     APPLICANT:  Jacobs, et al.
;     TITLE OF INVENTION:  L5 SHUTTLE PHASMIDS
;     NUMBER OF SEQUENCES:  1
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Amster, Rothstein & Ebenstein
;       STREET:  90 Park Avenue
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10016
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;       COMPUTER:  IBM PC Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  Word Processor (ASCII)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/075,904
;       FILING DATE:  May 11, 1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/247,901
;       FILING DATE:  May 23, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bogosian, Elizabeth A
;       REGISTRATION NUMBER:  39,911
```

```
;      REFERENCE/DOCKET NUMBER:  96700/475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50341
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  L5 shuttle phasmid sequence
;    HYPOTHETICAL:  No
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  L5 mycobacteriophage
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:  No. 5994137e
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1


  Query Match            62.5%;  Score 15;  DB 4;  Length 50341;
  Best Local Similarity  78.3%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGG 23
           || |  | | ||||| |||||||
Db   15790 CGGGTATGACTTGGGTGCCTTGG 15768


RESULT  15
US-08-673-269-9/c
; Sequence 9, Application US/08673269
; Patent No. 5756344
;  GENERAL INFORMATION:
;    APPLICANT:  ONDA, Haruo
;    APPLICANT:  OHKUBO, Shoichi
;    APPLICANT:  KOSAKA, Takuo
;    TITLE OF INVENTION:  DNA CODING FOR A HUMAN VASOCONSTRICTIVE
;    TITLE OF INVENTION:  PEPTIDE AND USE THEREOF
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,269
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/728169
;      FILING DATE:  10-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNIK, David S.
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  41109
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  99 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-673-269-9


  Query Match            60.8%;  Score 14.6;  DB 2;  Length 99;
  Best Local Similarity  81.0%;  Pred. No. 60;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 AGACTAATTTGGGTGCCTTGG 23
           || | ||  ||||||||||||
Db      26 AGCCGAAGGTGGGTGCCTTGG 6


Search completed: June 22, 2000, 18:04:06
Job time: 9713 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:34 ; Search time 2113.77 Seconds
(without alignments)
46.021 Million cell updates/sec

Title: US-09-362-485-23
Perfect score: 24
Sequence: 1 CGAGACTAATTTGGGTGCCTTGGC 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	18.2	75.8	257	27	AA023751	AA023751 mh81f01.r
	2	18.2	75.8	270	30	AA274532	AA274532 va91h10.r
	3	18.2	75.8	280	35	AA590517	AA590517 vi54b01.r
	4	18.2	75.8	361	79	AW259023	AW259023 um76d10.y
	5	18.2	75.8	363	47	AI527787	AI527787 uj29b04.y
	6	18.2	75.8	379	28	AA105433	AA105433 ml83g01.r
	7	18.2	75.8	457	109	AQ620908	AQ620908 HS_2178_B
	8	17.6	73.3	407	34	AA504826	AA504826 aa64d02.s
	9	17.6	73.3	493	28	AA097917	AA097917 mk17c10.r
c	10	17.6	73.3	694	94	AQ026314	AQ026314 l(3)j3D4
	11	17.4	72.5	568	61	AI861407	AI861407 614013E06
c	12	17.2	71.7	311	94	AQ067328	AQ067328 HS_2237_A
	13	17.2	71.7	328	29	AA148044	AA148044 zo53b12.r
	14	17.2	71.7	330	69	AV206428	AV206428 AV206428
	15	17.2	71.7	369	34	AA472398	AA472398 vh01e05.r
c	16	17.2	71.7	413	63	AI988386	AI988386 sd01c05.y
c	17	17.2	71.7	421	61	AI851219	AI851219 UI-M-BH0-
c	18	17.2	71.7	482	102	AQ144471	AQ144471 HS_3084_A
	19	17.2	71.7	484	44	AI304245	AI304245 ui64h09.y
	20	17.2	71.7	553	61	AI815009	AI815009 wk69h06.x
	21	16.8	70.0	349	40	AA987428	AA987428 oq93c10.s
	22	16.8	70.0	374	30	AA265270	AA265270 mx91c09.r
c	23	16.8	70.0	411	105	AQ278236	AQ278236 CITBI-E1-
c	24	16.8	70.0	441	108	AQ584558	AQ584558 RPCI-11-4
c	25	16.8	70.0	445	108	AQ584527	AQ584527 RPCI-11-4
	26	16.8	70.0	478	42	AI127372	AI127372 qb70e06.x
	27	16.8	70.0	525	107	AQ464800	AQ464800 HS_5105_B
c	28	16.8	70.0	543	108	AQ582902	AQ582902 RPCI-11-4
c	29	16.8	70.0	556	108	AQ524502	AQ524502 HS_5214_B
	30	16.8	70.0	593	83	FR0038368	AL125869 Fugu rubr
	31	16.8	70.0	598	107	AQ432378	AQ432378 HS_5079_A
c	32	16.6	69.2	188	27	C02017	C02017 HUMGS000475
	33	16.6	69.2	200	27	AA003890	AA003890 mg79h07.r
	34	16.6	69.2	216	48	AI551386	AI551386 vi54b01.y
	35	16.6	69.2	258	59	AI764980	AI764980 wh57c03.x
c	36	16.6	69.2	316	22	R21911	R21911 yh24a01.r1
c	37	16.6	69.2	322	24	H79957	H79957 yv79e09.r1
c	38	16.6	69.2	324	38	AA813816	AA813816 ai72d01.s
	39	16.6	69.2	334	23	H32314	H32314 EST107280 R
c	40	16.6	69.2	344	69	AW136550	AW136550 UI-H-BI1-
	41	16.6	69.2	357	27	AA009300	AA009300 mg96a07.r
	42	16.6	69.2	365	64	AL118144	AL118144 s8c10b62
	43	16.6	69.2	382	64	AL117905	AL117905 p9107b48
	44	16.6	69.2	387	30	AA222472	AA222472 mw20d04.r
	45	16.6	69.2	409	35	AA574941	AA574941 vm33f08.r

ALIGNMENTS

```
RESULT   1
AA023751
LOCUS       AA023751      257 bp    mRNA            EST       21-JAN-1997
DEFINITION  mh81f01.r1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA
            clone IMAGE:457369 5', mRNA sequence.
ACCESSION   AA023751
VERSION     AA023751.1  GI:1487666
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 257)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 8, 1995 this sequence version replaced gi:801201.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:274257
            Putative full length read
            vector to vector length is 273
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1. .257
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:457369"
                     /clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
                     /sex="unknown"
                     /tissue_type="placenta"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       70 a      54 c     105 g      28 t
ORIGIN


  Query Match           75.8%;  Score 18.2;  DB 27;  Length 257;
  Best Local Similarity 87.0%;  Pred. No. 51;
  Matches   20;  Conservati >    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GAGACTAATTTGGGTGCCTTGGC 24
          || ||| ||||||| ||||||||
Db    177 GATACTGATTTGGGCGCCTTGGC 199


RESULT   2
AA274532
LOCUS       AA274532      270 bp    mRNA            EST       31-MAR-1997
DEFINITION  va91h10.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
            IMAGE:746851 5', mRNA sequence.
ACCESSION   AA274532
VERSION     AA274532.1  GI:1915240
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 270)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802414.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
```

```
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:455835
            Putative full length read
            vector to vector length is 271
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .270
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:746851"
                      /clone_lib="Soares mouse 3NME12 5"
                      /sex="unknown"
                      /tissue_type="fetus"
                      /dev_stage="12.5dpc total fetus"
                      /lab_host="DH10B"
                      /note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Not I - oligo(dT) primer
                      [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT
                      3'], on total mouse RNA [provided by Minoru Ko, Wayne
                      State Univ.]; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library went through one round of normalization, and was
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       71 a     56 c    112 g     31 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 30;  Length 270;
  Best Local Similarity   87.0%;  Pred. No. 52;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGGC 24
           || ||| ||||||| ||||||||
Db     190 GATACTGATTTGGGCGCCTTGGC 212


RESULT   3
AA590517
LOCUS       AA590517      280 bp    mRNA            EST       16-SEP-1997
DEFINITION  vi54b01.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:907561 5', mRNA sequence.
ACCESSION   AA590517
VERSION     AA590517.1  GI:2405153
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 280)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393960.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:528225
            Putative full length read
            vector to vector length is 597
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 265.
FEATURES              Location/Qualifiers
     source           1. .280
                      /organism="Mus musculus"
                      /strain="C57BL6 x DBA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:907561"
                      /clone_lib="Beddington mouse embryonic region"
                      /sex="pooled"
                      /tissue_type="embryo"
                      /dev_stage="7.5dpc"
                      /lab_host="DH12S"
                      /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                      SalI; Site_2: NotI; Cloned unidirectionally. Primer:
                      Oligo dT. Gastrulating embryos were collected at 7.5dpc
                      from C57BL6 x DBA matings, excluding embryos that had
                      developed head folds and all extraembryonic tissues.
                      Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                      Referenced in Development 121, 2479-2489 (1995)"
BASE COUNT       63 a     61 c     92 g     64 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 35;  Length 280;
  Best Local Similarity   87.0%;  Pred. No. 52;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGG 23
           || |||| ||||||||||||| |
Db     178 CGGGACTGATTTGGGTGCCTTCG 200


RESULT   4
AW259023
LOCUS       AW259023      361 bp    mRNA            EST       23-DEC-1999
DEFINITION  um76d10.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2301139 5' similar to SW:IF2P_HUMAN O60841 PROBABLE
            TRANSLATION INITIATION FACTOR IF-2. [1] ;, mRNA sequence.
ACCESSION   AW259023
VERSION     AW259023.1  GI:6632004
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138169.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1009911
            Seq primer: custom primer used
            High quality sequence stop: 313.
FEATURES              Location/Qualifiers
     source           1. .361
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
```

```
                     /clone="IMAGE:2301139"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      111 a     71 c    133 g     46 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 79;  Length 361;
  Best Local Similarity   87.0%;  Pred. No. 53;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GAGACTAATTTGGGTGCCTTGGC 24
          || ||| ||||||| ||||||||
Db    194 GATACTGATTTGGGCGCCTTGGC 216


RESULT   5
AI527787
LOCUS       AI527787      363 bp    mRNA            EST       18-MAR-1999
DEFINITION  uj29b04.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1921327 5', mRNA sequence.
ACCESSION   AI527787
VERSION     AI527787.1  GI:4441922
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187749.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:977619
            Putative full length read
            vector to vector length is
            Seq primer: custom primer used
            High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1921327"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      113 a     72 c    136 g     42 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 47;  Length 363;
  Best Local Similarity   87.0%;  Pred. No. 53;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GAGACTAATTTGGGTGCCTTGGC 24
          || ||| ||||||| ||||||||
Db    174 GATACTGATTTGGGCGCCTTGGC 196


RESULT   6
AA105433
LOCUS       AA105433      379 bp    mRNA            EST       04-FEB-1997
DEFINITION  ml83g01.r1 Stratagene mouse kidney (#937315) Mus musculus cDNA
            clone IMAGE:518640 5', mRNA sequence.
ACCESSION   AA105433
VERSION     AA105433.1  GI:1655398
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1288998.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:312488
            Seq primer: -28ml3 rev1 ET from Amersham
            High quality sequence stop: 366.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Mus musculus"
                     /strain="C57/Bl6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:518640"
                     /clone_lib="Stratagene mouse kidney (#937315)"
                     /sex="females"
                     /tissue_type="kidney"
                     /dev_stage="4 weeks"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: pBluescript SK-; Site_1:
```

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 158 a 58 c 108 g 55 t
ORIGIN

Query Match 75.8%; Score 18.2; DB 28; Length 379;
Best Local Similarity 87.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGACTAATTTGGGTGCCTTGGC 24
|| ||| ||||||| ||||||||
Db 56 GATACTGATTTGGGCGCCTTGGC 78

RESULT 7
AQ620908
LOCUS AQ620908 457 bp DNA GSS 16-JUN-1999
DEFINITION HS_2178_B1_D02_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2178 Col=3 Row=H, genomic survey sequence.
ACCESSION AQ620908
VERSION AQ620908.1 GI:5083388
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 457)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT On Feb 19, 1999 this sequence version replaced gi:4128905.
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2178 row: H column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 457.
FEATURES Location/Qualifiers
source 1. .457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2178 Col=3 Row=H"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 150 a 40 c 108 g 154 t 5 others
ORIGIN

Query Match 75.8%; Score 18.2; DB 109; Length 457;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGACTAATTTGGGTGCCTTGGC 24
||||| |||||||||| |||||
Db 423 GAGACCAATTTGGGTGTATTGGC 445

RESULT 8
AA504826
LOCUS AA504826 407 bp mRNA EST 18-AUG-1997
DEFINITION aa64d02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:825699 3', mRNA sequence.
ACCESSION AA504826
VERSION AA504826.1 GI:2240986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 407)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 5, 1995 this sequence version replaced gi:798257.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers
source 1. .407
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:825699"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 95 a 106 c 107 g 99 t
ORIGIN

Query Match 73.3%; Score 17.6; DB 34; Length 407;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGAGACTAATTTGGGTGCCTTGGC 24
||||||| ||||||| ||||||
Db 155 CGAGACTCCCTTGGGTGGCTTGGC 178

RESULT 9
AA097917
LOCUS AA097917 493 bp mRNA EST 25-OCT-1996
DEFINITION mk17c10.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:493170 5', mRNA sequence.
ACCESSION AA097917

```
VERSION     AA097917.1  GI:1643529
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393883.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:296618
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:493170"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       95 a    129 c    122 g    147 t
ORIGIN
```

```
  Query Match             73.3%;  Score 17.6;  DB 28;  Length 493;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGGC 24
           || ||||| |||| ||||| ||||
Db     408 CGCGACTACTTTGTGTGCCCTGGC 431
```

```
RESULT  10
AQ026314/c
LOCUS       AQ026314      694 bp    DNA             GSS       30-JUN-1998
DEFINITION  1(3)j3D4 Drosophila melanogaster P lethal line Drosophila
            melanogaster genomic Sequence recovered from Both 5' and 3' ends of
            P element, genomic survey sequence.
ACCESSION   AQ026314
VERSION     AQ026314.1  GI:3266599
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Spradling,A.C., Stern,D., Beaton,A., Rehm,E.J., Laverty,T.,
            Mozden,N., Misra,S. and Rubin,G.M.
  TITLE     The BDGP gene disruption project: Single P element insertions
            mutating 30% of Drosophila autosomal genes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu

            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 118 in the 694 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.

            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster P lethal line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single P
                     transposable element insertion that is thought to cause
                     either lethality or sterility. The resultant fragment for
                     each strain was directly sequenced to determine the
                     genomic sequence at the site of insertion. Details of the
                     protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT      178 a    136 c    139 g    240 t      1 others
ORIGIN
```

```
  Query Match             73.3%;  Score 17.6;  DB 94;  Length 694;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCCTTGGC 24
           ||||| |||| || |||| |||||
Db     499 CGAGAGTAATGTGTGTGCATTGGC 476
```

```
RESULT  11
AI861407
LOCUS       AI861407      568 bp    mRNA            EST       19-JUL-1999
DEFINITION  614013E06.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI861407
VERSION     AI861407.1  GI:5525514
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1131208.
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
```

```
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614013  row: E  column: 06.
FEATURES              Location/Qualifiers
     source           1. .568
                      /organism="Zea mays"
                      /cultivar="W23"
                      /db_xref="taxon:4577"
                      /clone_lib="614 - root cDNA library from Walbot Lab"
                      /tissue_type="root"
                      /dev_stage="3-4 days old"
                      /lab_host="XLOLR"
                      /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                      EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                      Lab (LM)"
BASE COUNT      184 a    121 c    101 g    162 t
ORIGIN


  Query Match             72.5%;  Score 17.4;  DB 61;  Length 568;
  Best Local Similarity   94.7%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGAGACTAATTTGGGTGCC 19
           ||||||| ||||||||||||
Db     303 CGAGACAAATTTGGGTGCC 321


RESULT  12
AQ067328/c
LOCUS       AQ067328      311 bp    DNA              GSS       04-AUG-1998
DEFINITION  HS_2237_A1_D03_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2237 Col=5 Row=G, genomic survey
            sequence.
ACCESSION   AQ067328
VERSION     AQ067328.1  GI:3382846
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2237  row: G  column: 5
            Class: BAC ends
            High quality sequence stop: 311.
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=2237 Col=5 Row=G"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      107 a     64 c     47 g     93 t
ORIGIN


  Query Match             71.7%;  Score 17.2;  DB 94;  Length 311;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGG 23
           ||||||| |||| || ||||||
Db      71 GAGACTATTTTGCGTACCTTGG 50


RESULT  13
AA148044
LOCUS       AA148044      328 bp    mRNA             EST       05-DEC-1996
DEFINITION  zo53b12.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:590591 5', mRNA sequence.
ACCESSION   AA148044
VERSION     AA148044.1  GI:1717585
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394356.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This c_one is available royalty-free through LLNL ; contact the
            IMAGE Consortiu  (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 204.
FEATURES              Location/Qualifiers
     source           1. .328
                      /organism="Homo sapiens"
                      /db_xref="GDB:4621977"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:590591"
                      /clone_lib="Stratagene pancreas (#937208)"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pancreatic adenocarcinoma cell line. Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       81 a     73 c     96 g     77 t      1 others
ORIGIN


  Query Match             71.7%;  Score 17.2;  DB 29;  Length 328;
  Best Local Similarity   86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGG 23
           ||||||| ||||| |||| |||
Db     282 GAGACTACTTTGGTTGCCATGG 303


RESULT  14
AV206428
LOCUS       AV206428      330 bp    mRNA             EST       30-OCT-1999
```

```
DEFINITION  AV206428 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700086I19 3', mRNA sequence.
ACCESSION   AV206428
VERSION     AV206428.1  GI:6147281
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3136932.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1700086I19"
                     /clone_lib="RIKEN full-length enriched, adult male testis"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites. "
BASE COUNT      101 a     69 c     83 g     77 t
ORIGIN

  Query Match           71.7%;  Score 17.2;  DB 69;  Length 330;
  Best Local Similarity 86.4%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GAGACTAATTTGGGTGCCTTGG 23
           ||||||| |||| ||||||||
Db      83 GAGACTAGCTTGGATGCCTTGG 104

RESULT 15
AA472398
LOCUS       AA472398      369 bp    mRNA            EST       18-JUN-1997
DEFINITION  vh01e05.r1 Soar_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:874208 5', mRNA sequence.
ACCESSION   AA472398
VERSION     AA472398.1  GI:2200389
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404627.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:513688
            Putative full length read
            vector to vector length is
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:874208"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      103 a     58 c     89 g     119 t
ORIGIN
```

Query Match 71.7%; Score 17.2; DB 34; Length 369;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       3 AGACTAATTTGGGTGCCTTGGC 24
           |||||| ||||| ||| |||||
Db      67 AGACTAGTTTGGATGCTTTGGC 88
```

Search completed: June 22, 2000, 17:43:38
Job time: 20088 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:45 ; Search time 783.2 Seconds
(without alignments)
-19.873 Million cell updates/sec

Title: US-09-362-485-24
Perfect score: 16
Sequence: 1 ATTTGGGTGCCTTGGC 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882759 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		16	100.0	16	5	A87627	A87627 Sequence 24
2		16	100.0	24	5	A87626	A87626 Sequence 23
c 3		16	100.0	1194	2	MTU92472	U92472 Mycobacteri
c 4		16	100.0	1208	5	A89752	A89752 Sequence 9
c 5		16	100.0	1209	5	A87612	A87612 Sequence 9
c 6		16	100.0	1228	5	A87608	A87608 Sequence 5
c 7		16	100.0	1228	5	A89748	A89748 Sequence 5
c 8		16	100.0	1229	5	A87610	A87610 Sequence 7
c 9		16	100.0	1229	5	A89750	A89750 Sequence 7
c 10		16	100.0	1235	5	A87606	A87606 Sequence 3
c 11		16	100.0	1235	5	A87609	A87609 Sequence 6
c 12		16	100.0	1235	5	A87611	A87611 Sequence 8
c 13		16	100.0	1235	5	A89746	A89746 Sequence 3
c 14		16	100.0	1235	5	A89749	A89749 Sequence 6
c 15		16	100.0	1235	5	A89751	A89751 Sequence 8
c 16		16	100.0	1236	5	A87613	A87613 Sequence 10
c 17		16	100.0	1236	5	A89753	A89753 Sequence 10
c 18		16	100.0	1237	5	A87607	A87607 Sequence 4
c 19		16	100.0	1237	5	A89747	A89747 Sequence 4
c 20		16	100.0	1245	5	A87605	A87605 Sequence 2
c 21		16	100.0	1245	5	A89745	A89745 Sequence 2
c 22		16	100.0	1260	5	A87604	A87604 Sequence 1
c 23		16	100.0	1260	5	A89744	A89744 Sequence 1
c 24		16	100.0	2584	1	MTALADH	X63069 M.tuberculo
c 25		16	100.0	55414	1	MTV002	AL008967 Mycobacte
26		15	93.8	238	13	G06011	G06011 human STS W
27		15	93.8	697	13	G53965	G53965 SHGC-80918
28		15	93.8	2446	12	AF003836	AF003836 Mesocrice
29		15	93.8	64512	9	AP000255	AP000255 Homo sapi
30		15	93.8	95921	51	AC004141	AC004141 Homo sapi
31		15	93.8	100000	9	AP000213	AP000213 Homo sapi
32		15	93.8	109754	42	AC012023	AC012023 Homo sapi
33		15	93.8	111894	9	AP000135	AP000135 Homo sapi
34		15	93.8	120809	57	AC012090	AC012090 Homo sapi
35		15	93.8	149298	9	AP000031	AP000031 Homo sapi
36		15	93.8	150037	44	AC010534	AC010534 Homo sapi
c 37		15	93.8	158220	55	AC023786	AC023786 Homo sapi
c 38		15	93.8	160792	45	AC016171	AC016171 Homo sapi
39		15	93.8	171042	54	AC018846	AC018846 Homo sapi
40		15	93.8	171048	42	AC012050	AC012050 Homo sapi
c 41		15	93.8	181006	44	AC009231	AC009231 Homo sapi
c 42		15	93.8	183594	41	AC009302	AC009302 Homo sapi
43		15	93.8	204521	55	AC021959	AC021959 Homo sapi
c 44		14.4	90.0	840	7	CLJGACOIX	X59850 C.lacryma-j
45		14.4	90.0	918	13	CNS01FMM	AL142031 Anopheles

ALIGNMENTS

```
RESULT   1
A87627
LOCUS       A87627         16 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 24 from Patent WO9836089.
ACCESSION   A87627
VERSION     A87627.1  GI:6736267
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        1 a      3 c      6 g      6 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 16;
  Best Local Similarity  100.0%;  Pred. No. 46;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db      1 ATTTGGGTGCCTTGGC 16


RESULT   2
A87626
LOCUS       A87626         24 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 23 from Patent WO9836089.
ACCESSION   A87626
VERSION     A87626.1  GI:6736266
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      5 c      8 g      7 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 24;
  Best Local Similarity  100.0%;  Pred. No. 46;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db      9 ATTTGGGTGCCTTGGC 24


RESULT   3
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 2;  Length 1194;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT   4
A89752/c
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 1208;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT   5
A87612/c
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 1209;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT   6
A87608/c
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT   7
A89748/c
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 1228;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT   8
A87610/c
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match           100.0%;  Score 16;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789


RESULT   9
A89750/c
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
```

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 16;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789


RESULT  10
A87606/c
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN


  Query Match             100.0%;  Score 16;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790


RESULT  11
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 16;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789


RESULT  12
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 16;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789


RESULT  13
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

Query Match 100.0%; Score 16; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    805 ATTTGGGTGCCTTGGC 790
```

```
RESULT  14
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN
```

Query Match 100.0%; Score 16; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789
```

```
RESULT  15
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN
```

Query Match 100.0%; Score 16; DB 5; Length 1235;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATTTGGGTGCCTTGGC 16
          ||||||||||||||||
Db    804 ATTTGGGTGCCTTGGC 789
```

Search completed: June 22, 2000, 17:58:47
Job time: 10796 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:01:55 ; Search time 175.2 Seconds
(without alignments)
22.849 Million cell updates/sec

Title: US-09-362-485-24
Perfect score: 16
Sequence: 1 ATTTGGGTGCCTTGGC 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		16	100.0	16	1	V49525	Mycobacterium sp.
2		16	100.0	16	1	V49623	AlaDH derived olig
3		16	100.0	24	1	V49524	Mycobacterium sp.
4		16	100.0	24	1	V49622	AlaDH derived olig
c 5		16	100.0	1245	1	V49626	Mycobacterium tube
c 6		16	100.0	1260	1	V49510	Mycobacterium sp.
c 7		16	100.0	1260	1	V49625	Mycobacterium tube
8		14.4	90.0	3896	1	X05816	Mouse pheromone re
c 9		14	87.5	2061	1	V57439	Human protein kina
c 10		14	87.5	2327	1	V57440	Human protein kina
11		14	87.5	2717	1	N70621	Pho8 promoter and
12		14	87.5	3451	1	T28492	P. aeruginosa dete
13		14	87.5	3694	1	N70622	Pho8 promoter and
c 14		14	87.5	4325	1	Q81136	RPLA2-8 gene. Nove
15		13.4	83.8	296	1	T21886	Human gene signatu
16		13.4	83.8	317	1	T19152	Human gene signatu
c 17		13.4	83.8	423	1	Q12538	Pregnancy-specific
c 18		13.4	83.8	667	1	T78387	Coffee polyphenol
c 19		13.4	83.8	674	1	T78386	Coffee polyphenol
20		13.4	83.8	959	1	V15446	Human gene fragmen
c 21		13.4	83.8	1026	1	T79497	Meripilus giganteu
c 22		13.4	83.8	1097	1	V44860	Clone BG142_1 codi
23		13.4	83.8	1455	1	T00777	Bacillus sp. alkal
c 24		13.4	83.8	1573	1	Q05641	Clone hPS12 encodi
c 25		13.4	83.8	2077	1	Q30188	cDNA insert of pca
c 26		13.4	83.8	2077	1	Q83982	Cyclic-GMP stimula
c 27		13.4	83.8	2077	1	T51117	Heart calcium/calm
c 28		13.4	83.8	2077	1	T67203	Human heart 61 kDa
c 29		13.4	83.8	2077	1	V36161	cDNA encoding a hu
c 30		13.4	83.8	2077	1	V54766	Human 61 kDa CaM-P
c 31		13.4	83.8	2077	1	V48209	Human heart Ca2+/c
c 32		13.4	83.8	2693	1	Q30187	cDNA insert of pca
c 33		13.4	83.8	2693	1	Q83981	Cyclic-GMP stimula
c 34		13.4	83.8	2693	1	T51116	Hippocampus calciu
c 35		13.4	83.8	2693	1	T67202	Human brain 61 kDa
c 36		13.4	83.8	2693	1	V36160	cDNA encoding a hu
c 37		13.4	83.8	2693	1	V54765	Human 61 kDa CaM-P
c 38		13.4	83.8	2693	1	V48208	Human Ca2+/calmodu
c 39		13.4	83.8	3058	1	V34237	Human secreted pro
c 40		13.4	83.8	3438	1	Q05769	GALP3 Promoter. Re
41		13.4	83.8	3443	1	X19461	Human senescence f
42		13.4	83.8	3483	1	X19956	Human Tango-73 enc
43		13.4	83.8	4392	1	V37375	Streptococcus pneu
44		13.4	83.8	5009	1	Q21412	Carbonic anhydrase
45		13.4	83.8	7647	1	V15519	Hamster oral papil

ALIGNMENTS

RESULT 1
V49525
ID V49525 standard; DNA; 16 BP.
AC V49525;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R7.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 16 BP; 1 A; 3 C; 6 G; 6 T;

Query Match 100.0%; Score 16; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 1 ATTTGGGTGCCTTGGC 16

RESULT 2
V49623
ID V49623 standard; DNA; 16 BP.
AC V49623;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R7.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 16 BP; 1 A; 3 C; 6 G; 6 T;

Query Match 100.0%; Score 16; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 1 ATTTGGGTGCCTTGGC 16

RESULT 3
V49524
ID V49524 standard; DNA; 24 BP.
AC V49524;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R6.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 24 BP; 4 A; 5 C; 8 G; 7 T;

Query Match 100.0%; Score 16; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 9 ATTTGGGTGCCTTGGC 24

RESULT 4
V49622
ID V49622 standard; DNA; 24 BP.
AC V49622;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R6.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 24 BP; 4 A; 5 C; 8 G; 7 T;

Query Match 100.0%; Score 16; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 9 ATTTGGGTGCCTTGGC 24

RESULT 5
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for

CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 16; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 811 ATTTGGGTGCCTTGGC 796

RESULT 6
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 16; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGGC 16
||||||||||||||||
Db 826 ATTTGGGTGCCTTGGC 811

RESULT 7
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 16; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGG GCCTTGGC 16
||||||| ||||||||
Db 826 ATTTGGGTGCCTTGGC 811

RESULT 8
X05816
ID X05816 standard; cDNA; 3896 BP.
AC X05816;
DT 04-MAY-1999 (first entry)
DE Mouse pheromone receptor VR16 encoding cDNA.
KW Pheromone receptor; signal transduction; fertility; behaviour;
KW reproduction; rodent; insect; mouse; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 36. .266
FT /*tag= a
PN WO9900422-A1.
PD 07-JAN-1999.
PF 30-JUN-1998; U13680.
PR 30-JUN-1997; US-051284.
PA (HARD) HARVARD COLLEGE.
PI Buck L, Dulac C, Herrada G, Matsunami H;
DR WPI; 99-095684/08.
DR P-PSDB; W94916.
PT New isolated pheromone receptor polypeptides - used to develop
PT products for controlling fertility and behaviour in vertebrates and
PT invertebrates
PS Claim 18; Page 120-121; 308pp; English.
CC The invention relates to polynucleotide sequences encoding mammalian
CC pheromone receptor polypeptides. The polypeptides are expressed in
CC murine and rat vomeronasal organ. The products can be used for modifying
CC pheromone activity, e.g. for decreasing pheromone receptor mediated
CC signal transduction. They can be used for controlling fertility and
CC behaviour in vertebrates and invertebrates. Compositions comprising the
CC polypeptides are particularly useful in e.g. controlling fertility in
CC livestock and controlling reproduction in rodents or insects by
CC interrupting the normal behaviours of rodents or insects that result in
CC reproduction. The present sequence represents a cDNA encoding a mouse
CC pheromone receptor VR16. The cDNA is deposited under the Genbank
CC accession number AF011426.
SQ Sequence 3896 BP; 1182 A; 730 C; 746 G; 1238 T;

```
  Query Match             90.0%;  Score 14.4;  DB 1;  Length 3896;
  Best Local Similarity   93.8%;  Pred. No. 57;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATTTGGGTGCCTTGGC 16
           | ||||||||||||||
Db    1469 ACTTGGGTGCCTTGGC 1484


RESULT   9
V57439/c
ID   V57439 standard; cDNA; 2061 BP.
AC   V57439;
DT   21-DEC-1998  (first entry)
DE   Human protein kinase hYAK3-alpha cDNA.
KW   Protein kinase; hYAK-alpha; human; bone loss; osteoporosis;
KW   inflammation; acute respiratory distress syndrome; osteoarthritis;
KW   rheumatoid arthritis; inflammatory bowel disease; psoriasis;
KW   dermatitis; asthma; allergy; infection; HIV; immunodeficiency;
KW   cachexia; septic shock; pain; injury; cancer; testicular cancer;
KW   anorexia, bulimia; Parkinson's disease; cardiovascular disease;
KW   restenosis; atherosclerosis; heart failure; myocardial infarction;
KW   hypotension; hypertension; urinary retention; angina pectoris;
KW   ulcer; benign prostatic hypertrophy; psychosis; schizophrenia;
KW   anxiety; manic depression; delirium; severe mental retardat'on;
KW   dementia; Huntington's disease; Gilles de la Tourette syndr me;
KW   dyskinesia; diagnosis; therapy; vaccine; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             146. .1912
FT                   /*tag=  a
PN   EP-870825-A1.
PD   14-OCT-1998.
PF   05-MAR-1998; 301641.
PR   07-APR-1997; US-835170.
PR   05-MAR-1997; US-040618.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Creasy CL, Xie W;
DR   WPI; 98-523155/45.
DR   P-PSDB; W75790.
PT   New DNA encoding hYAK3 human protein kinase polypeptides - used to
PT   treat and diagnose e.g. osteoporosis, inflammatory diseases,
PT   inflammatory bowel disease, psoriasis, dermatitis, asthma,
PT   allergies, infections, septic shock, anorexia and cancer
PS   Claim 5; Fig 1; 23pp; English.
CC   This full-length cDNA clone codes for a novel human protein kinase,
CC   termed hYAK3-alpha (see W75790), which is expressed predominantly
CC   in skeletal muscle.  A partial clone was initially identified in a
CC   search of the Human Genome Sciences database on the basis of
CC   homology to yeast YAK1.  This partial clone was used to screen
CC   human testis and skeletal muscle cDNA libraries.  hYAK3-alpha cDNA
CC   was obtained from the skeletal muscle library, and hYAK3-beta cDNA
CC   (see V57440) from the testis library.  hYAK-alpha cDNA is 266
CC   nucleotides shorter that hYAK-beta cDNA and the 3' most 1844
CC   nucleotides are identical.  The differences between the 2 cDNAs may
CC   be due to alternative splicing.  The invention provides host cells
CC   and methods for producing hYAK3 polypeptides, as well as methods
CC   for treating subjects having need to enhance or inhibit hYAK3
CC   activity, for diagnosing a disease related to expression or
CC   activity of hYAK3, methods for identifying agonist and antagonist
CC   compounds, and polynucleotide probes and primers.  hYAK3
CC   polypeptides and polynucleotides can be used to treat and diagnose
CC   bone loss (e.g. osteoporosis), inflammatory diseases (e.g. ARDS),
CC   rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,
CC   psoriasis, dermatitis, asthma, allergies, infections (e.g.
CC   bacterial, fungal, protozoal and viral infections such as HIV),
CC   HIV-associated cachexia and other immunodeficiency disorders,
CC   septic shock, pain, injury, cancers (e.g. testicular cancer),
CC   anorexia, bulimia, Parkinson's disease, cardiovascular disease
CC   (e.g. restenosis, atherosclerosis, acute heart failure and
CC   myocardial infarction), hypotension, hypertension, urinary
CC   retention, angina pectoris, ulcers, benign prostatic hypertrophy
CC   and psychotic and neurological disorders (e.g. anxiety,
CC   schizophrenia, manic depression, delirium, dementia, severe mental
CC   retardation and dyskinesias such as Huntington's disease or Gilles
CC   de la Tourette syndrome).
SQ   Sequence   2061 BP;    574 A;    471 C;    514 G;    502 T;


  Query Match             87.5%;  Score 14;  DB 1;  Length 2061;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGG 15
           ||||||||||||||
Db     556 TTTGGGTGCCTTGG 543


RESULT  10
V57440/c
ID   V57440 standard; cDNA; 2327 BP.
AC   V57440;
DT   21-DEC-1998  (first entry)
DE   Human protein kinase hYAK3-beta cDNA.
KW   Protein kinase; hYAK-beta; human; bone loss; osteoporosis;
KW   inflammation; acute respiratory distress syndrome; osteoarthritis;
KW   rheumatoid arthritis; inflammatory bowel disease; psoriasis;
KW   dermatitis; asthma; allergy; infection; HIV; immunodeficiency;
KW   cachexia; septic shock; pain; injury; cancer; testicular cancer;
KW   anorexia, bulimia; Parkinson's disease; cardiovascular disease;
KW   restenosis; atherosclerosis; heart failure; myocardial infarction;
KW   hypotension; hypertension; urinary retention; angina pectoris;
KW   ulcer; benign prostatic hypertrophy; psychosis; schizophrenia;
KW   anxiety; manic depression; delirium; severe mental retardation;
KW   dementia; Huntington's disease; Gilles de la Tourette syndrome;
KW   dyskinesia; diagnosis; therapy; vaccine; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             472. .2178
FT                   /*tag=  a
PN   EP-870825-A1.
PD   14-OCT-1998.
PF   05-MAR-1998; 301641.
PR   07-APR-1997; US-835170.
PR   05-MAR-1997; US-040618.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Creasy CL, Xie W;
DR   WPI; 98-523155/45.
DR   P-PSDB; W75791.
PT   New DNA encoding hYAK3 human protein kinase polypeptides - used to
PT   treat and diagnose e.g. osteoporosis, inflammatory diseases,
PT   inflammatory bowel disease, psoriasis, dermatitis, asthma,
PT   allergies, infections, septic shock, anorexia and cancer
PS   Claim 5; Fig 2; 23pp; English.
CC   This full-length cDNA clone codes for a novel human protein kinase,
CC   termed hYAK3-beta (see W757901, which is expressed predominantly
CC   in testis.  A partial clone was initially identified in a search of
CC   the Human Genome Sciences database on the basis of homology to
CC   yeast YAK1.  This partial clone was used to screen human testis and
CC   skeletal muscle cDNA libraries.  hYAK3-beta cDNA was obtained from
CC   the testis cDNA library, and hYAK3-alpha cDNA (see V57439) from the
CC   skeletal muscle library.  hYAK-beta cDNA is 266 nucleotides longer
CC   than hYAK-alpha cDNA and the 3' most 1844 nucleotides are
CC   identical.  The differences between the 2 cDNAs may be due to
CC   alternative splicing.  The invention provides host cells and
CC   methods for producing hYAK3 polypeptides, as well as methods for
CC   treating subjects having need to enhance or inhibit hYAK3 activity,
CC   for diagnosing a disease related to expression or activity of
CC   hYAK3, methods for identifying agonist and antagonist compounds,
CC   and polynucleotide probes and primers.  hYAK3-alpha and -beta
CC   polypeptides and polynucleotides can be used to treat and diagnose
CC   bone loss (e.g. osteoporosis), inflammatory diseases (e.g. ARDS),
CC   rheumatoid arthritis, osteoarthritis, inflammatory bowel disease,
CC   psoriasis, dermatitis, asthma, allergies, infections (e.g.
```

```
CC   bacterial, fungal, protozoal and viral infections such as HIV),
CC   HIV-associated cachexia and other immunodeficiency disorders,
CC   septic shock, pain, injury, cancers (e.g. testicular cancer),
CC   anorexia, bulimia, Parkinson's disease, cardiovascular disease
CC   (e.g. restenosis, atherosclerosis, acute heart failure and
CC   myocardial infarction), hypotension, hypertension, urinary
CC   retention, angina pectoris, ulcers, benign prostatic hypertrophy
CC   and psychotic and neurological disorders (e.g. anxiety,
CC   schizophrenia, manic depression, delirium, dementia, severe mental
CC   retardation and dyskinesias such as Huntington's disease or Gilles
CC   de la Tourette syndrome).
SQ   Sequence   2327 BP;    660 A;    515 C;    573 G;    579 T;


  Query Match            87.5%;  Score 14;  DB 1;  Length 2327;
  Best Local Similarity  100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db    822 TTTGGGTGCCTTGG 809


RESULT  11
N70621
ID   N70621 standard; DNA; 2717 BP.
AC   N70621;
DT   30-APR-1991  (first entry)
DE   Pho8 promoter and gene in pBR322.
KW   Alkaline phosphatase; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   promoter        1. .989
FT                   /*tag=  a
FT                   /label= pho8 promoter
FT   cds             990. .2676
FT                   /*tag=  b
FT                   /label= pho8 gene
FT   misc_rna        2622. .2717
FT                   /*tag=  c
FT                   /label= part of pBR322
FT   tata_signal     865. .572
FT                   /*tag=  d
PN   J62151181-A.
PD   06-JUL-1987.
PF   23-DEC-1985; 287860.
PR   28-DEC-1984; JP-274456.
PR   09-APR-1985; JP-073493.
PR   10-SEP-1985; JP-198480.
PR   23-DEC-1985; JP-287860.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
DR   WPI; 87-225745/32.
PT   New recombinant DNA - contains a promoter gene which is cloned,
PT   for developing gene engineering prodn. of many polypeptide(s).
PS   Disclosure; Fig 5; 20pp; Japanese.
CC   The pho8 promoter can be used to construct vecoters for the prodn.
CC   of recombinant polypeptides.
CC   See also N70622.
SQ   Sequence   2717 BP;    825 A;    557 C;    657 G;    678 T;


  Query Match            87.5%;  Score 14;  DB 1;  Length 2717;
  Best Local Similarity  100.0%;  Pred. No. 90;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db   1898 TTTGGGTGCCTTGG 1911


RESULT  12
T28492
ID   T28492 standard; DNA; 3451 BP.
AC   T28492;
DT   01-APR-1997  (first entry)
DE   P. aeruginosa detection probe #3.
KW   Detection; probe; amplification primer; bacterial pathogen; pneumonia;
KW   Escherichia coli; Klebsiella pneumoniae; Pseudomonas aeruginosa;
KW   Proteus mirabilis; Streptococcus pneumoniae; Staphylococcus aureus;
KW   Staphylococcus epidermidis; Enterococcus faecalis; respiratory tract;
KW   Staphylococcus saprophyticus; Streptococcus pyogenes; urinary tract;
KW   Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;
KW   infection; intra-abdominal infection; skin infection;
KW   bacterial resistance; beta-lactam antibiotic; ds.
OS   Synthetic.
PN   WO9608582-A2.
PD   21-MAR-1996.
PF   12-SEP-1995; CA0528.
PR   12-SEP-1994; US-304732.
PA   (BERG/) BERGERON M G.
PA   (OUEL/) OUELLETTE M.
PA   (ROYP/) ROY P H.
PI   Bergeron MG,  Ouellette M,  Roy PH;
DR   WPI; 96-179953/18.
PT   Method for the detection of bacterial species using probes and
PT   primers  - allows detection and quantification of antibiotic
PT   resistant bacteria in patients, the environment and food
PS   Claim 34; Page 73-75; 216pp; English.
CC   The sequences given in T28490-503 represent probes which were used in
CC   the method of the invention for the detection of P. aeruginosa in a
CC   sample.  The method comprises using probes and/or amplification primers
CC   which are specific, ubiquitous and sensitive for determining the presence
CC   and/or amount of nucleic acids from selected bacterial species in any
CC   sample, where the bacterial nucleic acid comprises a selected target
CC   region hybridisable with the probes or primers.  The method comprises
CC   contacting the sample with the probes or primers and detecting the
CC   presence and/or amount of hybridised primers or amplification products
CC   as and indication of the presence and/or amount of the bacterial
CC   species.  This method may be used to detect commonly encountered
CC   bacterial pathogens, e.g. Escherichia coli, Klebsiella pneumoniae,
CC   Pseudomonas aeruginosa, Proteus mirabilis, Streptococcus pneumoniae,
CC   Staphylococcus aureus, Staphylococcus epidermidis, Enterococcus
CC   faecalis, Staphylococcus saprophyticus, Streptococcus pyogenes,
CC   Haemophilus influenzae and Moraxella catarrhalis.  These bacterial
CC   species are associated with approx. 90% of urinary tract infections and
CC   with a high percentage of other severe infections including
CC   septicaemia, meningitis, pneumonia, intra-abdominal infections, skin
CC   infections and other severe respiratory tract infections.  The method
CC   may also be used to evaluate a bacterial resistance to beta-lactam
CC   antibiotics.
SQ   Sequence   3451 BP;    717 A;    1061 C;    832 G;    841 T;


  Query Match            87.5%;  Score 14;  DB 1;  Length 3451;
  Best Local Similarity  100.0%;  Pred. No. 91;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 TTGGGTGCCTTGGC 16
          ||||||||||||||
Db   1822 TTGGGTGCCTTGGC 1839


RESULT  13
N70622
ID   N70622 standard; DNA; 3694 BP.
AC   N70622;
DT   30-APR-1991  (first entry)
DE   Pho8 promoter and gene.
KW   Alkaline phosphatase; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   promoter        1. .989
FT                   /*tag=  a
FT                   /label= pho8 promoter
FT   cds             990. .2690
FT                   /*tag=  b
```

```
FT                   /label= pho8 gene
FT   tata_signal     865. .872
FT                   /*tag=  c
PN   J62151181-A.
PD   06-JUL-1987.
PF   23-DEC-1985; 287860.
PR   28-DEC-1984; JP-274456.
PR   09-APR-1985; JP-073493.
PR   10-SEP-1985; JP-198480.
PR   23-DEC-1985; JP-287860.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
DR   WPI; 87-225745/32.
PT   New recombinant DNA - contains a promoter gene which is cloned,
PT   for developing gene engineering prodn. of many polypeptide(s).
PS   Disclosure; Fig 8; 20pp; Japanese.
CC   The pho8 promoter can be used to construct vectors for the prodn.
CC   of recombinant polypeptides.
CC   See also N70621.
SQ   Sequence   3694 BP;    1082 A;    729 C;    842 G;    1041 T;


  Query Match            87.5%;  Score 14;  DB 1;  Length 3694;
  Best Local Similarity  100.0%;  Pred. No. 91;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db   1898 TTTGGGTGCCTTGG 1911


RESULT  14
Q81136/c
ID   Q81136 standard; cDNA; 4325 BP.
AC   Q81136;
DT   15-AUG-1995  (first entry)
DE   RPLA2-8 gene.
KW   RPLA2-8; phospholipase A2; PLA2; Batten disease;
KW   neuronal ceroid lipfuscinosis; gene therapy; ss.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   cds             722. .1198
FT                   /*tag=  a
PN   WO9502328-A.
PD   26-JAN-1995.
PF   15-JUL-1994; U07926.
PR   15-JUL-1993; US-091941.
PR   26-JUL-1993; US-097354.
PA   (INCY-) INCYTE PHARM INC.
PA   (INDV ) UNIV INDIANA FOUND.
PI   Seilhamer JJ,  Tischfield JA;
DR   WPI; 95-067096/09.
DR   P-PSDB; R63044.
PT   Novel type III and IV low mol. wt. phospholipase A2 enzymes -
PT   from humans and rats, also nucleic acid sequences useful, e.g.
PT   for recombinant prodn. of enzymes, research into Batten's
PT   disease, etc.
PS   Disclosure; Page 57-60; 160pp; English.
CC   A human PLA2-encoding cDNA (Q81138) expressing HPLA2-10, was
CC   isolated from human brain RNA by RACE-PCR.  2 Rat PLA2 cDNAs,
CC   designated RPLA2-8 (Q81136) and RPLA2-10 (Q81137), were isolated from
CC   rat brain and heart cDNA libraries, respectively.  A partial human
CC   genomic counterpart to RPLA2-8, HPLA2-8 (Q81139), was also obtained.
CC   RPLA2-8 and HPLA2-8 have been designated type III PLA2, and RPLA2-10
CC   and HPLA2-10 as type IV.
SQ   Sequence   4325 BP;    1046 A;    1086 C;    1242 G;    951 T;


  Query Match            87.5%;  Score 14;  DB 1;  Length 4325;
  Best Local Similarity  100.0%;  Pred. No. 92;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db   1930 TTTGGGTGCCTTGG 1917


RESULT  15
T21886
ID   T21886 standard; cDNA to mRNA; 296 BP.
AC   T21886;
DT   14-AUG-1996  (first entry)
DE   Human gene signature HUMGS03428.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 992; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   296 BP;    60 A;    57 C;    97 G;    73 T;


  Query Match            83.8%;  Score 13.4;  DB 1;  Length 296;
  Best Local Similarity  93.3%;  Pred. No. 1.6e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGG 15
          || |||||||||||||
Db      2 ATCTGGGTGCCTTGG 16


Search completed: June 22, 2000, 18:01:58
Job time: 10130 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:50 ; Search time 2113.77 Seconds
(without alignments)
34.516 Million cell updates/sec

Title: US-09-362-485-21
Perfect score: 18
Sequence: 1 CGCGGCCGACATCATCGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	15.4	85.6	233	38	AA754395	AA754395 97MJ0326
	2	15.4	85.6	301	32	AA332158	AA332158 EST36085
	3	15.4	85.6	358	32	AA351471	AA351471 EST59231
	4	15.4	85.6	370	32	AA334266	AA334266 EST38455
	5	15.4	85.6	519	64	AW097386	AW097386 rs43c02.y
c	6	15.4	85.6	565	46	AI461152	AI461152 sa75g06.y
	7	15.4	85.6	591	49	AU056837	AU056837 AU056837
c	8	15.4	85.6	619	82	FR0013267	AL004517 F.rubripe
c	9	15.4	85.6	651	92	AQ956057	AQ956057 LERAH19TF
	10	15.4	85.6	688	92	AQ956058	AQ956058 LERAH19TR
	11	15.4	85.6	699	81	AV406190	AV406190 AV406190
	12	15.4	85.6	747	72	AW161035	AW161035 au78a04.y
	13	15.4	85.6	1478	94	AQ012082	AQ012082 483PLA2 C
	14	14.8	82.2	206	43	AI206334	AI206334 qr21g03.x
c	15	14.8	82.2	281	34	AA490408	AA490408 aa51b07.r
	16	14.8	82.2	300	35	C29149	C29149 C29149 Rice
	17	14.8	82.2	337	20	D15355	D15355 RICC0510A R
c	18	14.8	82.2	350	27	AA008309	AA008309 mg77c05.r
c	19	14.8	82.2	398	92	AQ967589	AQ967589 LERIT19TF
	20	14.8	82.2	410	79	AW274176	AW274176 xv27h11.x
c	21	14.8	82.2	411	74	AV394065	AV394065 AV394065
	22	14.8	82.2	435	42	AI131868	AI131868 uc37g03.r
	23	14.8	82.2	447	38	AA804041	AA804041 GM12621.5
	24	14.8	82.2	449	79	AW286179	AW286179 LG1_328_D
c	25	14.8	82.2	454	44	AI295414	AI295414 LP09033.3
	26	14.8	82.2	456	92	AQ963916	AQ963916 LERGP96TF
c	27	14.8	82.2	462	35	AA041067	AA041067 24333 CD4
	28	14.8	82.2	465	46	AI456657	AI456657 LD36751.5
c	29	14.8	82.2	541	36	AA621342	AA621342 af85c06.s
	30	14.8	82.2	545	42	AI107177	AI107177 GH06492.5
	31	14.8	82.2	554	63	AI979833	AI979833 pat.pk000
c	32	14.8	82.2	561	42	AI133900	AI133900 GH10923.3
c	33	14.8	82.2	566	40	AA990668	AA990668 LD34479.3
	34	14.8	82.2	566	109	AQ653674	AQ653674 Sheared D
c	35	14.8	82.2	587	43	AI238140	AI238140 GH14140.3
c	36	14.8	82.2	589	47	AI517647	AI517647 GH28687.5
	37	14.8	82.2	604	44	AI256967	AI256967 LP04077.5
	38	14.8	82.2	605	46	AI403859	AI403859 GH23389.5
c	39	14.8	82.2	615	44	AI295914	AI295914 LP09679.3
	40	14.8	82.2	616	40	AA978999	AA978999 LD33232.5
	41	14.8	82.2	618	44	AI259444	AI259444 LP02858.5
	42	14.8	82.2	620	44	AI259935	AI259935 LP03462.5
	43	14.8	82.2	626	44	AI296650	AI296650 LP10623.5
	44	14.8	82.2	629	47	AI515470	AI515470 LD47320.5
	45	14.8	82.2	631	45	AI388934	AI388934 GH19987.5

ALIGNMENTS

RESULT 1
AA754395
LOCUS AA754395 233 bp mRNA EST 20-JAN-1998
DEFINITION 97MJ0326 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 97MJ0326 similar to Pumpkin mRNA for glycolate oxidase, complete cds, mRNA sequence.
ACCESSION AA754395
VERSION AA754395.1 GI:2801101
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 233)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797800.
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
FEATURES Location/Qualifiers
source 1. .233
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="97MJ0326"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5'end with EcoRI and 3' end with Xho I site."
BASE COUNT 48 a 74 c 77 g 34 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 38; Length 233;
Best Local Similarity 94.1%; Pred. No. 8.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCG 17
||||||||| |||||||||
Db 152 CGCGGCCGGCATCATCG 168

RESULT 2
AA332158
LOCUS AA332158 301 bp mRNA EST 21-APR-1997
DEFINITION EST36085 Embryo, 8 week I Homo sapiens cDNA 5' end similar to collapsin response mediator protein 1, mRNA sequence.
ACCESSION AA332158
VERSION AA332158.1 GI:1984422
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 301)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., , Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407308.
Other_ESTs: THC173917
Contact: Kerlavage, AR

Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .301
/organism="Homo sapiens"
/db_xref="ATCC (inhost):133813"
/db_xref="taxon:9606"
/clone_lib="Embryo, 8 week I"
/dev_stage="embryo, 8 wks"
/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 61 a 94 c 91 g 52 t 3 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 32; Length 301;
Best Local Similarity 94.1%; Pred. No. 8.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCGGCCGACATCATCGC 18
|| |||||||||||||||
Db 126 GCAGCCGACATCATCGC 142

RESULT 3
AA351471
LOCUS AA351471 358 bp mRNA EST 21-APR-1997
DEFINITION EST59231 Infant brain Homo sapiens cDNA 5' end similar to collapsin response mediator protein 1, mRNA sequence.
ACCESSION AA351471
VERSION AA351471.1 GI:2003791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 358)
AUTHORS Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404776.
Other_ESTs: EST59230 THC173917
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .358
/organism="Homo sapiens"
/db_xref="ATCC (inhost):104427"
/db_xref="taxon:9606"
/clone_lib="Infant brain"
/sex="female"
/dev_stage="infant"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII; Site_2: NotI"

BASE COUNT 71 a 118 c 104 g 62 t 3 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 32; Length 358;
Best Local Similarity 94.1%; Pred. No. 8.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCGGCCGACATCATCGC 18
|| |||||||||||||||
Db 122 GCAGCCGACATCATCGC 138

RESULT 4
AA334266
LOCUS AA334266 370 bp mRNA EST 21-APR-1997
DEFINITION EST38455 Embryo, 9 week Homo sapiens cDNA 5' end similar to collapsin response mediator protein 1, mRNA sequence.
ACCESSION AA334266
VERSION AA334266.1 GI:1986509
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V , Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao ..., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.- , Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., , Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On May 9, 1995 this sequence version replaced gi:803018.
Other_ESTs: THC173917
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .370
/organism="Homo sapiens"
/db_xref="ATCC (inhost):135939"
/db_xref="taxon:9606"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 79 a 114 c 106 g 63 t 8 others
ORIGIN

```
  Query Match             85.6%;  Score 15.4;  DB 32;  Length 370;
  Best Local Similarity   94.1%;  Pred. No. 8.9e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATCGC 18
           || |||||||||||||||
Db      59 GCAGCCGACATCATCGC 75


RESULT   5
AW097386
LOCUS       AW097386      519 bp    mRNA            EST       19-OCT-1999
DEFINITION  rs43c02.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5'
            similar to TR:Q9Z2I7 Q9Z2I7 SV2 RELATED PROTEIN. ;, mRNA sequence.
ACCESSION   AW097386
VERSION     AW097386.1  GI:6067697
KEYWORDS    EST.
SOURCE      Pristionchus pacificus.
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria;
            Diplogasterida; Diplogasterina; Diplogasteroidea;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M.,
            Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S.,
            Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133849.
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
            Washington University Genome Sequencing Center
             Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for
            information about this clone.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Pristionchus pacificus"
                     /strain="PS 312"
                     /db_xref="taxon:54126"
                     /clone_lib="Sommer Pristionchus"
                     /sex="predominantly hermaphroditic"
                     /dev_stage="mixed stages (embryo to adult)"
                     /lab_host="not applicable (host cell line)"
                     /note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5'
                     EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a
                     XhoI - oligo(dT) primer. Double-stranded cDNA was ligated
                     to EcoRI adaptors digested with XhoI and cloned into XhoI
                     and EcoRI sites. Primary complexity of the library was 10
                     in the 7th. The library went through one round of
                     amplification."
BASE COUNT      108 a    147 c    129 g    131 t      4 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 64;  Length 519;
  Best Local Similarity   94.1%;  Pred. No. 9.5e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCG 17
           ||||| ||||||||||
Db     249 CGCGGTCGACATCATCG 265


RESULT   6
AI461152/c
LOCUS       AI461152      565 bp    mRNA            EST       01-DEC-1999
DEFINITION  sa75g06.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1004-5171 5' similar to SW:OSL3_ARATH P50700 OSMOTIN-LIKE
            PROTEIN OSM34 PRECURSOR. ;, mRNA sequence.
ACCESSION   AI461152
VERSION     AI461152.1  GI:4314033
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247248.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-5171"
                     /clone_lib="Gm-c1004"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Root cDNA. The mRNA was isolated from entire roots
                     of 8 day old 'Williams' seedlings which were propagated on
                     paper towels with distilled water. Stratagene's cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize the
                     cDNA. First- strand synthesis was performed with 5-methyl
                     dCTP, hence the ligated cDNA is hemimethylated.
                     Stratagene's first-strand synthesis primer was used
                     [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                     second-strand synthesis, the cDNA ends were 'polished'
                     with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                     and phosphorylated. The XhoI site within the first-strand
                     synthesis primer was restricted by digestion with XhoI;
                     all XhoI sites in the cDNA would be protected by their
                     hemimethylated status. The cDNA constructs were
                     size-fractionated with a 500bp cutoff, using GibcoBRL Life
                     Technologies' cDNA Size Fractionation column. The column
                     eluent was then ligated into Stratagene's pBluescript II
                     XR Predigested vector (pBluescript II SK(+) that had been
                     digested with EcoRI and XhoI, and phosphorylated). Both
                     the white and blue colonies appear to contain recombinant
                     plasmids with cDNA inserts. Blue colonies 9n=15) have been
                     sequenced, and possess putative cDNA inserts. This library
                     was constructed by Dr. Paul Keim & Virginia H. Coryell,
```

Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@nau.edu, virginia.coryell@nau.edu"
BASE COUNT 192 a 106 c 116 g 149 t 2 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 46; Length 565;
Best Local Similarity 94.1%; Pred. No. 9.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCG 17
||| ||||||||||||||
Db 563 CGCCGCCGACATCATCG 547

RESULT 7
AU056837
LOCUS AU056837 591 bp mRNA EST 29-APR-1999
DEFINITION AU056837 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA clone S20918_1A, mRNA sequence.
ACCESSION AU056837
VERSION AU056837.1 GI:4715721
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 591)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from mature leaf
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187289.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .591
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S20918_1A"
/clone_lib="Oryza sativa mature leaf Nipponbare"
/tissue_type="mature leaf"
BASE COUNT 134 a 151 c 185 g 119 t 2 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 49; Length 591;
Best Local Similarity 94.1%; Pred. No. 9.8e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCG 17
|||||||| ||||||||
Db 511 CGCGGCCGGCATCATCG 527

RESULT 8
FR0013267/c
LOCUS FR0013267 619 bp DNA GSS 18-SEP-1997
DEFINITION F.rubripes GSS sequence, clone 113017bA2, genomic survey sequence.
ACCESSION AL004517
VERSION AL004517.1 GI:2450087
KEYWORDS GSS; genome survey sequence.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE 1 (bases 1 to 619)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .619
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 113017"
/clone="113017bA2"
BASE COUNT 159 a 162 c 137 g 120 t 41 others
ORIGIN

Query Match 85.6%; Score 15.4; DB 82; Length 619;
Best Local Similarity 88.9%; Pred. No. 9.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
| |||||||||||||||
Db 263 CANGGCCGACATCATCGC 246

RESULT 9
AQ956057/c
LOCUS AQ956057 651 bp DNA GSS 28-JAN-2000
DEFINITION LERAH19TF LERA Arabidopsis thaliana genomic clone LERAH19, genomic survey sequence.
ACCESSION AQ956057
VERSION AQ956057.1 GI:6783646
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 651)
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL Unpublished (2000)
COMMENT On Dec 15, 1999 this sequence version replaced gi:4575012.
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For addtional information, see http://www.tigr.org/tdb/at/at.html
Seq primer: TF
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .651
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"

```
                     /clone="LERAH19"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT       120 a    185 c    209 g    137 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 92;  Length 651;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATCGC 18
           ||||||||| |||||||
Db     615 GCGGCCGACGTCATCGC 599


RESULT  10
AQ956058
LOCUS       AQ956058      688 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERAH19TR LERA Arabidopsis thaliana genomic clone LERAH19, genomic
            survey sequence.
ACCESSION   AQ956058
VERSION     AQ956058.1  GI:6783647
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     On Dec 15, 1999 this sequence version replaced gi:4575013.
            Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAH19"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT       138 a    236 c    189 g    125 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 92;  Length 688;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATCGC 18
           ||||||||| |||||||
Db     339 GCGGCCGACGTCATCGC 355


RESULT  11
AV406190
LOCUS       AV406190      699 bp    mRNA            EST       06-FEB-2000
DEFINITION  AV406190 Bombyx mori wing disk C108 5th-instar day-3 larva Bombyx
            mori cDNA clone wdV30773 T3, mRNA sequence.
ACCESSION   AV406190
VERSION     AV406190.1  GI:6910278
KEYWORDS    EST.
SOURCE      domestic silkworm.
  ORGANISM  Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
  TITLE     Bombyx mori cDNA
  JOURNAL   Unpublished (2000)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5422412.
            Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
            Email: kmita@nirs.go.jp
            method:uni-directional, sequence direction:sequenced from T3 primer
            (5' -> 3')
             Project='Silkworm Genome Program in MAFF, and Research for the
            Future Program in JSPS'. see 'SilkBase',
            <http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Bombyx mori"
                     /strain="C108"
                     /db_xref="taxon:7091"
                     /clone="wdV30773"
                     /clone_lib="Bombyx mori wing disk C108 5th-instar day-3
                     larva"
                     /sex="female/male mixed"
                     /tissue_type="wing disk"
                     /dev_stage="5th-instar day-3 larva"
BASE COUNT       181 a    171 c    205 g    142 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 81;  Length 699;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCG 17
           ||||||||| |||||||||
Db     543 CGCGGCCGTCATCATCG 559


RESULT  12
AW161035
LOCUS       AW161035      747 bp    mRNA            EST       09-NOV-1999
DEFINITION  au78a04.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
            IMAGE:2782350 5' similar to SW:DPY1_HUMAN Q14194
            DIHYDROPYRIMIDINASE RELATED PROTEIN-1 ;, mRNA sequence.
ACCESSION   AW161035
VERSION     AW161035.1  GI:6300068
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 6, 1998 this sequence version replaced gi:3114949.
            Other_ESTs: au78a04.x1
            Contact: Wilson RK
            Washington University School of Medicine
```

```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 435.
FEATURES              Location/Qualifiers
     source           1. .747
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2782350"
                      /clone_lib="Schneider fetal brain 00004"
                      /sex="male"
                      /tissue_type="frontal lobe"
                      /dev_stage="5 months post-conception"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pBluescript SK (Stratagene);
                      Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                      prepared from human fetal brain tissue.  5' and 3'
                      adaptors were used in cloning as follows: 5' adaptor
                      sequence:
                      5'-GAGAGAGAGAAGAGCTCAAGGATCCTTAATTAAATTAATCCCCCCCCCCCCCC-3'
                      and 3' adaptor sequence:
                      5'-GAGAGAGAGACTCGAGTTTTTTTTTTTTTTTTT-3'.  The library was
                      size-selected for >0.5 kb inserts and has an average
                      insert size estimated at 1.2 kb.  This library was
                      constructed using the CAP-trapper method for full-length
                      enrichment and has not undergone amplification.    Library
                      was constructed by Dr. Claudio Schneider (LNCIB-Area
                      Science Park,  Trieste, Italy). "
BASE COUNT       176 a    217 c    217 g    137 t
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 72;  Length 747;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATCGC 18
           || |||||||||||||||
Db      34 GCAGCCGACATCATCGC 50


RESULT  13
AQ012082
LOCUS       AQ012082     1478 bp    DNA              GSS       04-JUN-1998
DEFINITION  483PLA2 Cosmid library of chromosome II Rhodobacter sphaeroides
            genomic clone 483PLA2, genomic survey sequence.
ACCESSION   AQ012082
VERSION     AQ012082.1  GI:3177037
KEYWORDS    GSS.
SOURCE      Rhodobacter sphaeroides.
  ORGANISM  Rhodobacter sphaeroides
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodobacter.
REFERENCE   1  (bases 1 to 1478)
  AUTHORS   Choudhary,M., Mackenzie,C., Mouncey,N., Weinstock,G.M. and
            Kaplan,S.
  TITLE     RsGDB, the Rhodobacter sphaeroides Genome Database
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Choudhary, M.
            Department of Microbiology and Molecular Genetics
            University of Texas Medical School
            6431 fannin Street, Houston, TX 77030, USA
            Tel: 713 500 5437
            Fax: 713 500 5499
            Email: madhu@utmmg.med.uth.tmc.edu
            Seq primer: pLA2
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .1478
                      /organism="Rhodobacter sphaeroides"
                      /strain="2.4.1T"
                      /db_xref="taxon:1063"
                      /clone="483PLA2"
                      /clone_lib="Cosmid library of chromosome II"
                      /lab_host="E. coli S17-1"
                      /note="Vector: pLA2917"
BASE COUNT       244 a    491 c    500 g    242 t      1 others
ORIGIN


  Query Match             85.6%;  Score 15.4;  DB 94;  Length 1478;
  Best Local Similarity   94.1%;  Pred. No. 1.2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCG 17
           ||||||||||||| |||
Db     433 CGCGGCCGACATCCTCG 449


RESULT  14
AI206334
LOCUS       AI206334      206 bp    mRNA            EST       16-OCT-1998
DEFINITION  qr21g03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:1941556 3'
            similar to contains MER22.t1 MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI206334
VERSION     AI206334.1  GI:3765006
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .206
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1941556"
                      /clone_lib="NCI_CGAP_GC6"
                      /tissue_type="pooled germ cell tumors"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Plasmid DNA from the normalized library
                      NCI_CGAP_GC4 was prepared, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from a pool of 5,000 clones made
                      from the same library (cloneIDs 1257096-1258631,
                      1469064-1470983, and 1475592-1476743). Subtraction by
                      Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       15 a     68 c     95 g     28 t
ORIGIN
```

```
  Query Match            82.2%;  Score 14.8;  DB 43;  Length 206;
  Best Local Similarity  88.9%;  Pred. No. 1.5e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           |||||||||| |||| ||
Db      31 CGCGGCCGACTTCATGGC 48


RESULT  15
AA490408/c
LOCUS       AA490408      281 bp    mRNA            EST       15-AUG-1997
DEFINITION  aa51b07.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824437 5'
            similar to SW:P044_RAT P38718 0-44 PROTEIN. ;, mRNA sequence.
ACCESSION   AA490408
VERSION     AA490408.1  GI:2219581
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393174.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 172.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:824437"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       56 a     84 c     95 g     45 t      1 others
ORIGIN


  Query Match            82.2%;  Score 14.8;  DB 34;  Length 281;
  Best Local Similarity  88.9%;  Pred. No. 1.6e+03;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           |||||||||||||  |||
Db     182 CGCGGCCGACATCGCCGC 165


Search completed: June 22, 2000, 17:43:30
Job time: 20080 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:40 ; Search time 783.2 Seconds
(without alignments)
-24.841 Million cell updates/sec

Title: US-09-362-485-22
Perfect score: 20
Sequence: 1 GGCCGACATCATCGCTTCCC 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		20	100.0	20	5	A87625	A87625 Sequence 22
2	c	20	100.0	1235	5	A87609	A87609 Sequence 6
3	c	20	100.0	1235	5	A87611	A87611 Sequence 8
4	c	20	100.0	1235	5	A89749	A89749 Sequence 6
5	c	20	100.0	1235	5	A89751	A89751 Sequence 8
6	c	20	100.0	1236	5	A87613	A87613 Sequence 10
7	c	20	100.0	1236	5	A89753	A89753 Sequence 10
8	c	20	100.0	1237	5	A87607	A87607 Sequence 4
9	c	20	100.0	1237	5	A89747	A89747 Sequence 4
10	c	20	100.0	1245	5	A87605	A87605 Sequence 2
11	c	20	100.0	1245	5	A89745	A89745 Sequence 2
12	c	20	100.0	1260	5	A87604	A87604 Sequence 1
13	c	20	100.0	1260	5	A89744	A89744 Sequence 1
14	c	20	100.0	2584	1	MTALADH	X63069 M.tuberculo
15	c	20	100.0	56414	1	MTV002	AL008967 Mycobacte
16	c	17	85.0	1235	5	A87606	A87606 Sequence 3
17	c	17	85.0	1235	5	A89746	A89746 Sequence 3
18		16.8	84.0	56414	1	MTV002	AL008967 Mycobacte
19		16.4	82.0	32050	1	AB017508	AB017508 Bacillus
20		15.8	79.0	939	12	AF020346	AF020346 Rattus no
21	c	15.8	79.0	1470	1	TFERBCR	D11141 T.ferrooxid
22	c	15.8	79.0	1508	7	ASNMETA	L37524 Aspergillus
23		15.8	79.0	2171	12	AF115249	AF115249 Rattus no
24	c	15.8	79.0	2318	1	YEFYUG1	Z35485 Y.enterocol
25	c	15.8	79.0	2318	1	YEFYUG2	Z35486 Y.enterocol
26	c	15.8	79.0	2318	1	YEFYUG3	Z35487 Y.enterocol
27	c	15.8	79.0	2318	1	YEFYUG4	Z35496 Y.enterocol
28	c	15.8	79.0	2395	1	YSFYUAG	Z29675 Y.enterocol
29		15.8	79.0	3699	16	AF020336	AF020336 Rice ragg
30	c	15.8	79.0	3812	2	AF160727	AF160727 Salmonell
31		15.8	79.0	4438	1	SYOAPCE	D63788 Synechococc
32	c	15.8	79.0	4500	43	AC017709	AC017709 Drosophil
33	c	15.8	79.0	5334	2	AF004408	AF004408 Rhizobium
34	c	15.8	79.0	7727	34	AGU42214	U42214 Anopheles g
35	c	15.8	79.0	31685	34	CELF16F9	U67956 Caenorhabdi
36	c	15.8	79.0	39399	1	MSGB1970CS	L78815 Mycobacteri
37		15.8	79.0	40479	34	CELK12C11	AF043701 Caenorhab
38		15.8	79.0	79554	50	T20M3	AC009999 Sequence
39		15.8	79.0	136163	33	AC004873	AC004873 Homo sapi
40	c	15.8	79.0	143537	42	AC009463	AC009463 Trypanoso
41		15.8	79.0	154086	54	AC021891	AC021891 Oryza sat
42	c	15.8	79.0	158900	55	AC008368	AC008368 Trypanoso
43	c	15.8	79.0	165678	33	HS50O24	AL034380 Homo sapi
44		15.8	79.0	167559	52	AC022197	AC022197 Homo sapi
45	c	15.8	79.0	195327	53	AC019228	AC019228 Homo sapi

ALIGNMENTS

```
RESULT   1
A87625
LOCUS       A87625         20 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 22 from Patent WO9836089.
ACCESSION   A87625
VERSION     A87625.1  GI:6736265
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        3 a      9 c      4 g      4 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 5;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db      1 GGCCGACATCATCGCTTCCC 20


RESULT   2
A87609/c
LOCUS       A87609       1235 bp    DNA             P T       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match           100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1235 GGCCGACATCATCGCTTCCC 1216


RESULT   3
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1235 GGCCGACATCATCGCTTCCC 1216


RESULT   4
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match           100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1235 GGCCGACATCATCGCTTCCC 1216


RESULT   5
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1235 GGCCGACATCATCGCTTCCC 1216


RESULT   6
A87613/c
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1236 GGCCGACATCATCGCTTCCC 1217


RESULT   7
A89753/c
LOCUS       A89753       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9832862.
ACCESSION   A89753
VERSION     A89753.1  GI:6738287
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    395 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1236;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1236 GGCCGACATCATCGCTTCCC 1217


RESULT   8
A87607/c
LOCUS       A87607       1237 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9836089.
ACCESSION   A87607
VERSION     A87607.1  GI:6736247
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    394 c    386 g    221 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1237;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1236 GGCCGACATCATCGCTTCCC 1217


RESULT   9
A89747/c
LOCUS       A89747       1237 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9832862.
ACCESSION   A89747
VERSION     A89747.1  GI:6738281
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    394 c    386 g    221 t
ORIGIN


  Query Match             100.0%;  Score 20;  DB 5;  Length 1237;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1236 GGCCGACATCATCGCTTCCC 1217


RESULT  10
A87605/c
LOCUS       A87605       1245 bp    DNA             PAT       22-JAN-2000
```

```
DEFINITION  Sequence 2 from Patent WO9836089.
ACCESSION   A87605
VERSION     A87605.1  GI:6736245
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 5;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1242 GGCCGACATCATCGCTTCCC 1223


RESULT  11
A89745/c
LOCUS       A89745       1245 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9832862.
ACCESSION   A89745
VERSION     A89745.1  GI:6738279
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 5;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1242 GGCCGACATCATCGCTTCCC 1223


RESULT  12
A87604/c
LOCUS       A87604       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9836089.
ACCESSION   A87604
VERSION     A87604.1  GI:6736244
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 5;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1257 GGCCGACATCATCGCTTCCC 1238


RESULT  13
A89744/c
LOCUS       A89744       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9832862.
ACCESSION   A89744
VERSION     A89744.1  GI:6738278
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 5;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1257 GGCCGACATCATCGCTTCCC 1238


RESULT  14
MTALADH/c
LOCUS       MTALADH      2584 bp    DNA             BCT       01-DEC-1993
DEFINITION  M.tuberculosis gene for L-alanine dehydrogenase.
ACCESSION   X63069 S36765
VERSION     X63069.1  GI:44565
KEYWORDS    alanine dehydrogenase; extracellular; intracellular.
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 2584)
  AUTHORS   Andersen,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1991) A.B. Andersen, Statens Seruminstitut,
            Artillerivej 5, DK-2300 Copenhagen S, DENMARK
REFERENCE   2  (bases 1 to 2584)
  AUTHORS   Andersen,A.B., Andersen,P. and Ljungqvist,L.
  TITLE     Structure and function of a 40,000-molecular-weight protein antigen
            of Mycobacterium tuberculosis
```

```
  JOURNAL   Infect. Immun. 60 (6), 2317-2323 (1992)
  MEDLINE   92267644
FEATURES             Location/Qualifiers
     source          1. .2584
                     /organism="Mycobacterium tuberculosis"
                     /strain="Erdman"
                     /isolate="TMC 107, lot 9A-2"
                     /db_xref="taxon:1773"
                     /clone_lib="lambda gt11"
                     /clone="lambda AA67"
     CDS             76. .1197
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="alanine  dehydrogenase"
                     /protein_id="CAA44791.1"
                     /db_xref="GI:44566"
                     /db_xref="SWISS-PROT:P30234"
                     /translation="MRVGIPTETKNNEFQFRVAITPAGVAELTRRGHEVLIQAGAGEG
                     SAITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAAS
                     RACTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGV
                     LMGGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIH
                     TRYSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGC
                     FEGSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAA
                     CRSNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      463 a    871 c    819 g    431 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 1;  Length 2584;
  Best Local Similarity  100.0%;  Pred. No. 25;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db   1257 GGCCGACATCATCGCTTCCC 1238


RESULT  15
MTV002/c
LOCUS       MTV002      56414 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
ACCESSION   AL008967 AL123456
VERSION     AL008967.1  GI:3261491
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 56414)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 56414)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2624256.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .56414
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .>56414
                     /note="fragment designated v002. Does not represent a
                     physical clone"
     source          <1. .233
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y154"
     gene            complement(3. .527)
                     /gene="recX"
     CDS             complement(3. .527)
                     /gene="recX"
                     /note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar
                     to eg. RECX_PSEAE P37860 regulatory protein recx from
                     Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161
                     z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
                     overlap. Overlaps and extends CDS from ovelapping cosmid
                     MTCY154.16c"
                     /codon_start=1
                     /transl_table=11
                     /product="recX"
                     /protein_id="CAA15532.1"
                     /db_xref="GI:2624257"
                     /db_xref="SPTREMBL:O33280"
                     /translation="MTVSCPPPSTSEREEQARALCLRLLTARSRTRAELAGQLAKRGY
                     PEDIGNRVLDRLAAVGLVDDTDFAEQWVQSRRANAAKSKRALAAELHAKGVDDDVITT
                     VLGGIDAGAERGRAEKLVRARLRREVLIDDGTDEARVSRRLVAMLARRGYGQTLACEV
                     VIAELAAERERRRV"
     gene            complement(493. .2865)
                     /gene="recA"
     CDS             complement(493. .2865)
                     /gene="recA"
                     /note="Rv2737c, (MTV002.02c), len: 790 aa. recA,
                     identicalto RECA_MYCTU P26345 reca protein (790 aa).
                     Contains self-splicing protein element (intein) from 2294
                     to 969 (c), similar to intein II from TR:E332317
                     (EMBL:Y13030) DNA-directedDNA polymerase (EC 2.7.7.7) from
                     Thermococcus sp. (1829 aa), fasta scores; opt: 81 z-score:
                     235.2 E(): 6e-06, 24.6%identity in 183 aa overlap.
                     Contains PS00017 ATP/GTP-bindingsite motif A (P-loop),
                     PS00321 recA signature, and PS00881Protein splicing
                     signature. See Davis et al, (1992) Cell71(2):201-210"
                     /codon_start=1
                     /transl_table=11
                     /product="recA"
                     /protein_id="CAA15533.1"
                     /db_xref="GI:2624258"
                     /db_xref="SWISS-PROT:P26345"
                     /translation="MTQTPDREKALELAVAQIEKSYGKGSVMRLGDEARQPISVIPTG
                     SIALDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGVAAFIDAEHALDP
```

```
                DYAKKLGVDTDSLLVSQPDTGEQALEIADMLIRSGALDIVVIDSVAALVPRAELEGEM
                GDSHVGLQARLMSQALRKMTGALNNSGTTAIFINQLRDKIGVMFGSPETTTGGKALKF
                YASVRMDVRRVETLKDGTNAVGNRTRVKVVKNKCLAEGTRIFDPVTGTTHRIEDVVDG
                RKPIHVVAAAKDGTLHARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRA
                AGELRKGDRVAQPRRFDGFGDSAPIPADHARLLGYLIGDGRDGWVGGKTPINFINVQR
                ALIDDVTRIAATLGCAAHPQGRISLAIAHRPGERNGVADLCQQAGIYGKLAWEKTIPN
                WFFEPDIAADIVGNLLFGLFESDGWVSREQTGALRVGYTTTSEQLAHQIHWLLLRFGV
                GSTVRDYDPTQKRPSIVNGRRIQSKRQVFEVRISGMDNVTAFAESVPMWGPRGAALIQ
                AIPEATQGRRRGSQATYLAAEMTDAVLNYLDERGVTAQEAAAMIGVASGDPRGGMKQV
                LGASRLRRDRVQALADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTL
                VAEGVVVHNCSPPFKQAEFDILYGKGISREGSLIDMGVDQGLIRKSGAWFTYEGEQLG
                QGKENARNFLVENADVADEIEKKIKEKLGIGAVVTDDPSNDGVLPAPVDF"
misc_feature    complement(790. .2112)
                /gene="recA"
                /note="recA intein"
misc_feature    complement(790. .807)
                /gene="recA"
                /note="PS00881 Protein splicing signature"
misc_feature    complement(2197. .2223)
                /gene="recA"
                /note="PS00321 recA signature"
misc_feature    complement(2644. .2667)
                /gene="recA"
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
RBS             complement(2873. .2877)
                /note="possible RBS upstream of Rv2737c"
gene            complement(3247. .3453)
                /gene="Rv2738c"
CDS             complement(3247. .3453)
                /gene="Rv2738c"
                /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown,
                N-terminus is highly similar to the N-terminus of the
                upstream ORFMTV002.07c (78.4% identity in 37 aa overlap);
                also similarto AL020958|SC4H8_5 Streptomyces coelicolor
                cosmid 4H8 (64aa)  opt:  185 z-score: 283.5 E(): 2.9e-08;
                39.7% identityin 63 aa overlap'
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2738c"
                /protein_id="CAA15534.1"
                /db_xref="GI:2624260"
                /db_xref="SPTREMBL:O33281"
                /translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHVLTGFDGRSAAQAI
                EDGVEPRDVWRALCADFDVPHDRW"
gene            complement(3464. .4630)
                /gene="Rv2739c"
CDS             complement(3464. .4630)
                /gene="Rv2739c"
                /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable
                transferase, similar to eg. TR:Q51560 (EMBL:L28170)
                Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
                fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
                identity in 425 aaoverlap. Equivalent to Mycobacterium
                leprae protein MLCB33.02c (392 aa); fasta scores
                gp|Z94723|MLCB33_2 Mycobacteriumleprae cosmid B33 opt:
                2112 z-score: 2364.5 E(): 0; 80.9%identity in 388 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2739c"
                /protein_id="CAA15535.1"
                /db_xref="GI:2624261"
                /db_xref="SPTREMBL:O33282"
                /translation="MRVAVVAGPDPGHSFPAIALCQRFRAAADTPTLFTGVEWLEAAR
                AAGIDAVELDGLAATDRDLDAGARIHRRAAQMAVLNVPRLRALEPELVVSDVITACGG
                MAAELLGIPWVELNPHPLYLPSKGLPPIGSGLAAGTGIRGRLRDATMRALTGRSWRAG
                LRQRAAVRVEIGLPARDPGPLRRLIATLPALEVPRPDWPAEAVVVGPLHFEPTDRVLA
                IPAGTGPVVVVAPSTALTGTAGLTEVALQSLTPGETVPSGSRLVVSRLSGADLTVPPW
                AVAGLGSQAELLTRADLVICGGGHGMVAKTLLAGVPMVVVPGGGDQWEIANRVVRQGS
                AVLIRPLTADALVAAVNEVLSSPRFREAARRAAASVAGAADPVRVCHDALALAG"
gene            4674. .5123
                /gene="Rv2740"
CDS             4674. .5123
                /gene="Rv2740"
                /note="Rv2740, (MTV002.05), len: 149 aa. unknown, similar
                to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
                MLCB33.03 (B2235_F3_140) (178 aa), fasta scores; opt: 498
                z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2740"
                /protein_id="CAA15536.1"
                /db_xref="GI:2624262"
                /db_xref="SPTREMBL:O33283"
                /translation="MAELTETSPETPETTEAIRAVEAFLNALQNEDFDTVDAALGDDL
                VYENVGFSRIRGGRRTATLLRRMQGRVGFEVKIHRIGADGAAVLTERTDALIIGPLRV
                QFWVCGVFEVDDGRITLWRDYFDVYDMFKGLLRGLVALVVPSLKATL"
gene            5355. .6932
                /gene="PE_PGRS"
CDS             5355. .6932
                /gene="PE_PGRS"
                /note="Rv2741, (MTV002.06), len: 525 aa; Member of M.
                tuberculosis PE_ PGRS subfamily, similar to many eg.
```

Query Match 100.0%; Score 20; DB 1; Length 56414;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db  39436 GGCCGACATCATCGCTTCCC 39417
```

Search completed: June 22, 2000, 17:58:43
Job time: 10792 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:01:44 ; Search time 175.2 Seconds
(without alignments)
28.561 Million cell updates/sec

Title: US-09-362-485-22
Perfect score: 20
Sequence: 1 GGCCGACATCATCGCTTCCC 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		20	100.0	20	1	V49523	Mycobacterium sp.
2		20	100.0	20	1	V49621	AlaDH derived olig
3	c	20	100.0	1245	1	V49626	Mycobacterium tube
4	c	20	100.0	1260	1	V49510	Mycobacterium sp.
5	c	20	100.0	1260	1	V49625	Mycobacterium tube
6		15.2	76.0	1487	1	X00715	Human secreted pro
7		15.2	76.0	1525	1	X00679	Human secreted pro
8		15.2	76.0	1575	1	T30361	Bovine p57 coding
9		15.2	76.0	8501	1	T12907	Novel expression p
10	c	15.2	76.0	9974	1	X20573	Polynucleotide seq
11	c	15.2	76.0	50341	1	V22674	DNA sequence of a
12	c	15.2	76.0	52297	1	T51411	Mycobacteriophage
13	c	15.2	76.0	52298	1	Q47357	L5 mycobacteriopha
14		15.2	76.0	110000	1	V30458_2	Continuation (3 of
15	c	15.2	76.0	110000	1	V30458_3	Continuation (4 of
16		15.2	76.0	110000	1	V30459_2	Continuation (3 of
17	c	15.2	76.0	110000	1	V30459_3	Continuation (4 of
18		15	75.0	18	1	V49522	Mycobacterium sp.
19		15	75.0	18	1	V49620	AlaDH derived olig
20		14.8	74.0	888	1	V31126	Aspergillus oryzae
21	c	14.8	74.0	1539	1	V60753	HIV-1 strain YBF30
22	c	14.8	74.0	2916	1	T15957	Cycloisomaltooligo
23	c	14.8	74.0	9183	1	V60751	HIV-1 strain YBF30
24		14.8	74.0	134588	1	Q04888	Total base sequenc
25	c	14.8	74.0	134525	1	Q04525	Total base sequenc
26		14.4	72.0	523	1	Q56098	Probe for Cryptosp
27	c	14.4	72.0	523	1	Q56099	Probe for Cryptosp
28		14.2	71.0	78	1	V38796	Probe used to clon
29		14.2	71.0	81	1	V62685	Soybean Kunitz typ
30		14.2	71.0	84	1	V36956	Soybean Kunitz try
31		14.2	71.0	84	1	V36521	Soybean Kunitz try
32		14.2	71.0	166	1	V88436	EST clone GF209. N
33	c	14.2	71.0	189	1	T33833	Open reading frame
34		14.2	71.0	261	1	T05846	Upstream 5' flanki
35		14.2	71.0	384	1	T66441	P14K polypeptide D
36	c	14.2	71.0	519	1	Q82824	GDF-7 C-terminal r
37	c	14.2	71.0	564	1	T33823	Open reading frame
38	c	14.2	71.0	959	1	X21103	Polynucleotide seq
39	c	14.2	71.0	1203	1	Q96223	Murine mV1 protein
40	c	14.2	71.0	1203	1	T90396	Murine BMP-12 homo
41	c	14.2	71.0	1253	1	N71054	Optimized Escheric
42		14.2	71.0	1410	1	X07013	Bovine papillomavi
43	c	14.2	71.0	1800	1	X14317	H. pylori GHPO 126
44	c	14.2	71.0	2051	1	T33810	RAGE-4 gene. RAGE
45	c	14.2	71.0	2058	1	N71053	Escherichia coli p

ALIGNMENTS

RESULT 1
V49523
ID V49523 standard; DNA; 20 BP.
AC V49523;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R5.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 20 BP; 3 A; 9 C; 4 G; 4 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||||||||||||||||
Db      1 GGCCGACATCATCGCTTCCC 20
```

RESULT 2
V49621
ID V49621 standard; DNA; 20 BP.
AC V49621;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R5.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 20 BP; 3 A; 9 C; 4 G; 4 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||||||||||||||||
Db 1 GGCCGACATCATCGCTTCCC 20

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 20; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||||||||||||||||
Db 1242 GGCCGACATCATCGCTTCCC 1223

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 20; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||||||||||||||||
Db 1257 GGCCGACATCATCGCTTCCC 1238

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 20; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||||||||||||||||
Db 1257 GGCCGACATCATCGCTTCCC 1238

RESULT 6
X00715
ID X00715 standard; DNA; 1487 BP.
AC X00715;
DT 25-MAR-1999 (first entry)
DE Human secreted protein gene 69 clone HLDBQ19.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9842738-A1.
PD 01-OCT-1998.
PF 19-MAR-1998; U05311.
PR 30-MAY-1997; US-050937.
PR 21-MAR-1997; US-041276.
PR 21-MAR-1997; US-041277.
PR 21-MAR-1997; US-041281.
PR 21-MAR-1997; US-042344.
PR 30-MAY-1997; US-048069.
PR 30-MAY-1997; US-048094.
PR 30-MAY-1997; US-048095.
PR 30-MAY-1997; US-048096.
PR 30-MAY-1997; US-048099.
PR 30-MAY-1997; US-048131.
PR 30-MAY-1997; US-048135.
PR 30-MAY-1997; US-048154.
PR 30-MAY-1997; US-048160.
PR 30-MAY-1997; US-048186.
PR 30-MAY-1997; US-048187.
PR 30-MAY-1997; US-048188.
PR 30-MAY-1997; US-048350.
PR 30-MAY-1997; US-048351.
PR 30-MAY-1997; US-048352.
PR 30-MAY-1997; US-048355.
PR 05-AUG-1997; US-054804.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA,
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,
PI Rosen CA, Ruben SM, Shi Y, Young P;
DR WPI; 99-070066/06.
DR P-PSDB; W67911.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1; Page 263-264; 385pp; English.
CC This sequence represents a nucleic acid molecule which encodes a secreted
CC human protein. The gene number, and the clone it is derived from, are
CC detailed in the descriptor line. The gene can be used to generate fusion
CC proteins by linking to the gene to a human immunoglobulin Fc portion
CC (e.g. X00602) for increasing the stability of the fused protein as
CC compared to the human protein only.
CC The invention relates to 87 novel genes and their fragments (nucleic acid
CC sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 87 polynucleotides, based on
CC which tissues they are most highly expressed in (see X00611 for described
CC uses).
SQ Sequence 1487 BP; 312 A; 447 C; 445 G; 278 T;

Query Match 76.0%; Score 15.2; DB 1; Length 1487;
Best Local Similarity 85.0%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||| |||||| | |||
Db 680 GGCCGACCTCATCGGTCCCC 699

RESULT 7
X00679
ID X00679 standard; DNA; 1525 BP.
AC X00679;
DT 25-MAR-1999 (first entry)
DE Human secreted protein gene 69 clone HLDBQ19.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN WO9842738-A1.
PD 01-OCT-1998.
PF 19-MAR-1998; U05311.
PR 30-MAY-1997; US-050937.
PR 21-MAR-1997; US-041276.
PR 21-MAR-1997; US-041277.
PR 21-MAR-1997; US-041281.
PR 21-MAR-1997; US-042344.
PR 30-MAY-1997; US-048069.
PR 30-MAY-1997; US-048094.
PR 30-MAY-1997; US-048095.
PR 30-MAY-1997; US-048096.
PR 30-MAY-1997; US-048099.
PR 30-MAY-1997; US-048131.
PR 30-MAY-1997; US-048135.
PR 30-MAY-1997; US-048154.
PR 30-MAY-1997; US-048160.
PR 30-MAY-1997; US-048186.
PR 30-MAY-1997; US-048187.
PR 30-MAY-1997; US-048188.
PR 30-MAY-1997; US-048350.
PR 30-MAY-1997; US-048351.
PR 30-MAY-1997; US-048352.
PR 30-MAY-1997; US-048355.
PR 05-AUG-1997; US-054804.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA,
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,
PI Rosen CA, Ruben SM, Shi Y, Young P;
DR WPI; 99-070066/06.
DR P-PSDB; W67875.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders

```
PS   Claim 1; Page 229; 385pp; English.
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein.  The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line.  The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. X00602) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 87 novel genes and their fragments (nucleic acid
CC   sequences: X00611-X00724; amino acid sequences W67807-W68004) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 87 polynucleotides, based on
CC   which tissues they are most highly expressed in (see X00611 for described
CC   uses).
SQ   Sequence   1525 BP;    284 A;    502 C;    458 G;    276 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 1525;
  Best Local Similarity  85.0%;  Pred. No. 68;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||||| |||||| | |||
Db    680 GGCCGACCTCATCGGTCCCC 699


RESULT   8
T30361
ID   T30361 standard; cDNA; 1575 BP.
AC   T30361;
DT   26-NOV-1996  (first entry)
DE   Bovine p57 coding sequence.
KW   WD40 repeated structure region; bovine p57 protein; dimer;
KW   anti-cancer; anti-HIV agent; actin; cell movement; ds.
OS   Bos taurus.
FH   Key             Location/Qualifiers
FT   cds             93. .1478
FT                   /*tag= a
FT                   /product= p57
PN   J08119996-A.
PD   14-MAY-1996.
PF   21-OCT-1994; 282743.
PR   21-OCT-1994; JP-282743.
PA   (NISB ) JAPAN TOBACCO INC.
DR   WPI; 96-283507/29.
DR   P-PSDB; R98342.
PT   A new protein, p57, comprising WD40 repeat region - used for
PT   development of anti-cancer and anti-HIV agents
PS   Example 2; Page 27-29; 51pp; Japanese.
CC   This sequence encodes the bovine p57 protein.  p57 contains a leucine
CC   rich C-terminal peptide which comprises a leucine residue after each
CC   7 amino acids and a  WD40 repeated structure region containing five
CC   WD40 regions.  p57 forms a dimer.  It can be used in the development
CC   of an anti-cancer agent and an anti-HIV agent.  It can also combine
CC   with actin in the control of cell movement.
SQ   Sequence   1575 BP;    304 A;    482 C;    492 G;    297 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 1575;
  Best Local Similarity  85.0%;  Pred. No. 69;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          |||||||| | |||| ||||
Db   1494 GGCCGACAGCCTCGCCTCCC 1513


RESULT   9
T12907
ID   T12907 standard; cDNA; 8501 BP.
AC   T12907;
DT   05-NOV-1996  (first entry)
DE   Novel expression plasmid pXL2435.
KW   Expression vector; T7 phage; gene 10; promoter; heterologous; circular;
KW   transcriptional terminator; plasmid stability region; RP4; maintenance;
KW   plasmid loss; antibiotic resistance; selection; pharmaceutical;
KW   agricultural; biocatalysis; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    6038. .8499
FT                   /*tag= a
FT                   /note= "contains sequence derived from par region
FT                   of plasmid RP4"
PN   WO9608572-A1.
PD   21-MAR-1996.
PF   14-SEP-1995; F01178.
PR   16-SEP-1994; FR-011049.
PA   (RHON ) RHONE POULENC RORER SA.
PI   Cameron B,  Crouzet J;
DR   WPI; 96-179946/18.
PT   Expression plasmids contg. phage T7 promoter - and stabilising par
PT   region from plasmid RP4
PS   Claim 11; Page 14-19; 35pp; French.
CC   This is the nucleotide sequence of a novel expression plasmid derived
CC   from the pET series of expression vectors.  The plasmid retains the T7
CC   phage gene 10 promoter and transcriptional terminator for expression of
CC   heterologous proteins e.g. basic fibroblast growth factor. The novelty of
CC   the plasmid derives from the inclusion of the plasmid stability region of
CC   plasmid RP4.  This region allows maintenance of the plasmid and prevents
CC   plasmid loss without the need for antibiotic resistance genes for
CC   selection and maintenance. The RP4 par region comprises the genes parA,
CC   B, C, D and E. They are placed in the plasmid under control of the lacO
CC   operator and lacI(q) repressor. The novel plasmid can be used to express
CC   proteins of pharmaceutical, agricultural or biocatalytic interest without
CC   the use of antibiotics in the culture medium which, when present in trace
CC   amounts in final preparations, could lead to undesired side effects or
CC   antibiotic resistant microorganisms flourishing.
SQ   Sequence   8501 BP;    1862 A;    2422 C;    2426 G;    1791 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 8501;
  Best Local Similarity  85.0%;  Pred. No. 81;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          ||||| | ||||||| ||||
Db   7806 GGCCGCCCTCATCGCGTCCC 7825


RESULT  10
X20573/c
ID   X20573 standard; DNA; 9974 BP.
AC   X20573;
DT   05-MAY-1999  (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum.
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI; 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1; Page 582-587; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
```

CC production of biosynthetic products such as enzymes.
SQ Sequence 9974 BP; 2342 A; 2344 C; 2757 G; 2523 T;

Query Match 76.0%; Score 15.2; DB 1; Length 9974;
Best Local Similarity 85.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
|| ||||| ||||||||||
Db 4009 GGGCGACAATATCGCTTCCC 3990

RESULT 11
V22674/c
ID V22674 standard; DNA; 50341 BP.
AC V22674;
DT 15-JUL-1998 (first entry)
DE DNA sequence of a L5 shuttle phasmid designated phAE41.
KW L5 shuttle phasmid; phAE41; Escherichia coli bacteriophage cosmid;
KW pYUB328; mycobacteria; mutation; vaccine; treatment;
KW mycobacterial infection; ss.
OS Chimeric - Bacteriophage L5.
OS Chimeric - Bacteriophage cosmid pYUB328.
PN US5750384-A.
PD 12-MAY-1998.
PF 23-MAY-1994; 247901.
PR 23-MAY-1994; US-247901.
PR 07-FEB-1992; US-833431.
PR 29-APR-1993; US-057531.
PA (UYPI-) UNIV PITTSBURGH.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PI Bardarov S, Hatfull GF, Jacobs WR, McAdam R;
DR WPI; 98-296771/26.
PT Mycobacterial shuttle plasmids comprising mycobacteriophage L5 with
PT lambda cosmid insert - useful for delivery of foreign DNA into
PT mycobacteria and for generating mutations
PS Claim 7; Columns 5-28; 42pp; English.
CC The present sequence represents a L5 shuttle phasmid designated phAE41.
CC It was prepared by inserting an Escherichia coli bacteriophage cosmid,
CC pYUB328, into a non-essential region of an L5 mycobacteriophage genome.
CC Foreign DNA such as reporter genes (e.g. luciferase), transposons
CC (e.g. IS1096) and mycobacterial inhibiter genes can also be inserted
CC into the phasmid. The phasmids can be used to introduce foreign DNA
CC into mycobacteria. The L5 shuttle phasmid can be used to introduce
CC mutations that can be used to study the mechanisms of mycobacteria and
CC to develop vaccines and drugs for treating mycobacterial infections.
SQ Sequence 50341 BP; 9641 A; 15673 C; 15276 G; 9750 T;

Query Match 76.0%; Score 15.2; DB 1; Length 50341;
Best Local Similarity 85.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
|||||| |||||| |||| |
Db 8578 GGCCGAGATCATCTCTTCTC 8559

RESULT 12
T51411/c
ID T51411 standard; DNA; 52297 BP.
AC T51411;
DT 23-APR-1997 (first entry)
DE Mycobacteriophage L5 genome sequence.
KW Reporter mycobacteriophage; Mycobacterium tuberculosis; infection;
KW drug susceptibility; diagnosis; ss.
OS Mycobacteriophage L5.
PN WO9425572-A1.
PD 10-NOV-1994.
PF 29-APR-1994; U04788.
PR 29-APR-1993; US-057531.
PA (UYPI-) UNIV PITTSBURGH.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PI Bloom BR, Hatfull GF, Jacobs WR;
DR WPI; 95-022252/03.
PT Reporter myco-bacteriophages - useful for rapid diagnosis of
PT myco-bacterial infection and assessment of drug susceptibilities to
PT mycobacterial strains.
PS Disclosure; Page 103-137; 192pp; English.
CC The entire genome sequence of mycobacteriophage L5 is given in
CC T51411. L5 is a temperate virus with broad host range among
CC mycobacteria. Transcriptional promoters and reporter genes (esp.
CC luciferase or beta-galactosidase) may be introduced into the genome
CC sequence to provide mycobacterial species-specific reporter
CC mycobacteriophages, e.g. phGS1 (ATCC 75454) and phGS5 (ATCC 75453).
CC These are useful in the rapid diagnosis of mycobacterial infection
CC and for the assessment of drug susceptibilities of mycobacteria in
CC clinical samples. Specifically, a reporter mycobacteriophage can
CC be used to diagnose tuberculosis and to assess the drug
CC susceptbilities of various strains of Mycobacterium tuberculosis.
SQ Sequence 52297 BP; 9633 A; 16532 C; 16025 G; 10107 T;

Query Match 76.0%; Score 15.2; DB 1; Length 52297;
Best Local Similarity 85.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
|||||| |||||| |||| |
Db 8475 GGCCGAGATCATCTCTTCTC 8456

RESULT 13
Q47357/c
ID Q47357 standard; DNA; 52298 BP.
AC Q47357;
DT 07-JUN-1994 (first entry)
DE L5 mycobacteriophage DNA.
KW Genome; L5; mycobacteriophage; temperate virus; mycobacteria; phage;
KW phage g; lysogeny; integration; chromosome; attP; attachment site;
KW 62-61-60 gene; left arm; integrase; integration-proficient vector;
KW reporter gene; diagnosis; right arm; tuberculosis; drug resistance;
KW mycobacterial species-specific reporter mycobacteriophage; ss.
OS L5 mycobacteriophage.
PN WO9316172-A.
PD 19-AUG-1993.
PF 02-FEB-1993; U00913.
PR 07-FEB-1992; US-833431.
PA (UYPI-) UNIV PITTSBURGH.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PI Bloom BR, Hatfull GF, Jacobs WR;
DR WPI; 93-272877/34.
PT Mycobacterial species-specific reporter myco-bacteriophages -
PT used to detect infection by or drug resistance of mycobacteria
PS Disclosure; Page 11-25; 65pp; English.
CC This sequence represents the DNA sequence of the entire genome of
CC the L5 mycobacteriophage. L5 is a temperate virus with a broad host
CC range among mycobacteria, and is morphologically similar to the
CC family of phages that includes phage g and contain a linear double
CC stranded genome with cohesive ends. During the establishment of
CC lysogeny, the L5 genome becomes integrated into the mycobacterial
CC chromosome at the attP attachment site. DNA analysis has indicated
CC that the L5 genome is organised into a right and left arm with the
CC attachment site and integrase at the center of the genome. The L5
CC genome may be used for the production of integration-proficient
CC vectors for mycobacteria. Part of the L5 genome is not essential for
CC mycobateriophage growth. It has been demonstrated that all or part of
CC the gene 62-61-60 can be deleted without effecting the life cycle of
CC the L5 phage. This region is therefore suitable for insertion of
CC reporter genes. Mycobacterial species-specific reporter mycobacterio-
CC phages produced in this way may be used to diagnose mycobacterial
CC disease, preferably tuberculosis, by incubating with a sample and
CC detecting the reporter gene product. They may also be used for
CC assessing drug resistance of a mycobacterial strain by adding the
CC reporter vector to the strain, then adding the drug and detecting

```
CC   the formation of any gene product.
SQ   Sequence   52298 BP;    9631 A;    16531 C;    16022 G;    10106 T;


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 52298;
  Best Local Similarity  85.0%;  Pred. No. 95;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          |||||| |||||| |||| |
Db   8475 GGCCGAGATCATCTCTTCTC 8456


RESULT  14
V30458_2
Continuation (3 of 6) of V30458 from base 200001 (Rhizobium species plasmid pNGR234a. )
WP Sequence split into 6 fragments  LOCUS V30458 Accession V30458
WP     Fragment Name      Begin     End
WP     V30458_0               1    110000
WP     V30458_1          100001    210000
WP     V30458_2          200001    310000
WP     V30458_3          300001    410000
WP     V30458_4          400001    510000
WP     V30458_5          500001    534720


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 110000;
  Best Local Similarity  85.0%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          |||||||||   ||||||||
Db  70475 GGCCGACATGGCCGCTTCCC 70494


RESULT  15
V30458_3/c
Continuation (4 of 6) of V30458 from base 300001 (Rhizobium species plasmid pNGR234a. )
WP Sequence split into 6 fragments  LOCUS V30458 Accession V30458
WP     Fragment Name      Begin     End
WP     V30458_0               1    110000
WP     V30458_1          100001    210000
WP     V30458_2          200001    310000
WP     V30458_3          300001    410000
WP     V30458_4          400001    510000
WP     V30458_5          500001    534720


  Query Match            76.0%;  Score 15.2;  DB 1;  Length 110000;
  Best Local Similarity  85.0%;  Pred. No. 1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCCC 20
          |||||||||||| | | |||
Db  46647 GGCCGACATCATAGATACCC 46628


Search completed: June 22, 2000, 18:01:53
Job time: 10125 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:03:58 ; Search time 123.79 Seconds
(without alignments)
21.001 Million cell updates/sec

Title: US-09-362-485-22
Perfect score: 20
Sequence: 1 GGCCGACATCATCGCTTCCC 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	15.2	76.0	50341	2	US-08-247-901C-1	Sequence 1, Appli
c	2	15.2	76.0	50341	4	US-09-075-904-1	Sequence 1, Appli
c	3	14.2	71.0	189	4	US-08-530-569B-20	Sequence 20, Appl
	4	14.2	71.0	261	4	US-08-867-030B-2	Sequence 2, Appli
	5	14.2	71.0	261	6	PCT-US95-06119-2	Sequence 2, Appli
	6	14.2	71.0	384	2	US-08-726-136-21	Sequence 21, Appl
c	7	14.2	71.0	519	4	US-08-581-528A-5	Sequence 5, Appli
c	8	14.2	71.0	519	6	PCT-US94-07799-5	Sequence 5, Appli
c	9	14.2	71.0	564	4	US-08-530-569B-18	Sequence 18, Appl
c	10	14.2	71.0	1203	1	US-08-362-670B-29	Sequence 29, Appl
c	11	14.2	71.0	1203	5	US-08-333-576C-29	Sequence 29, Appl
c	12	14.2	71.0	1203	6	PCT-US94-14030A-29	Sequence 29, Appl
c	13	14.2	71.0	2051	4	US-08-530-569B-13	Sequence 13, Appl
c	14	14.2	71.0	2369	3	US-08-883-534-4	Sequence 4, Appli
c	15	14.2	71.0	2369	5	US-09-204-764-4	Sequence 4, Appli
	16	14.2	71.0	2574	3	US-08-677-734A-8	Sequence 8, Appli
	17	14.2	71.0	3323	3	US-07-624-299-14	Sequence 14, Appl
	18	14.2	71.0	4131	1	US-08-485-588-4	Sequence 4, Appli
	19	14.2	71.0	4131	2	US-08-484-565-4	Sequence 4, Appli
	20	14.2	71.0	4131	3	US-08-480-751-4	Sequence 4, Appli
	21	14.2	71.0	4131	4	US-08-943-986-4	Sequence 4, Appli
	22	14.2	71.0	4131	5	US-08-353-784-4	Sequence 4, Appli
c	23	14.2	71.0	11219	2	US-07-642-734C-1	Sequence 1, Appli
	24	14.2	71.0	49377	2	US-08-764-233A-1	Sequence 1, Appli
c	25	14	70.0	1249	4	US-08-933-750C-62	Sequence 62, Appl
	26	13.8	69.0	36	3	US-08-468-413-5	Sequence 5, Appli
	27	13.8	69.0	36	6	PCT-US95-07169-5	Sequence 5, Appli
c	28	13.8	69.0	840	7	5432081-6	Patent No. 5432081
c	29	13.8	69.0	1215	3	US-08-758-621-3	Sequence 3, Appli
	30	13.8	69.0	1257	4	US-08-776-210-2	Sequence 2, Appli
c	31	13.8	69.0	1488	7	5268463-6	Patent No. 5268463
	32	13.8	69.0	2563	1	US-08-423-441-1	Sequence 1, Appli
	33	13.8	69.0	3211	4	US-08-574-959A-8	Sequence 8, Appli
	34	13.8	69.0	3901	4	US-08-574-959A-6	Sequence 6, Appli
	35	13.8	69.0	5056	3	US-08-793-126-2	Sequence 2, Appli
c	36	13.8	69.0	7742	3	US-08-882-704A-4	Sequence 4, Appli
c	37	13.6	68.0	96	1	US-08-303-124-2	Sequence 2, Appli
c	38	13.6	68.0	96	1	US-08-204-729-2	Sequence 2, Appli
c	39	13.6	68.0	96	3	US-08-480-697B-2	Sequence 2, Appli
	40	13.6	68.0	100	1	US-08-303-124-1	Sequence 1, Appli
	41	13.6	68.0	100	1	US-08-204-729-1	Sequence 1, Appli
	42	13.6	68.0	100	3	US-08-480-697B-1	Sequence 1, Appli
	43	13.6	68.0	112	1	US-08-303-124-23	Sequence 23, Appl
	44	13.6	68.0	112	3	US-08-480-697B-23	Sequence 23, Appl
	45	13.6	68.0	236	2	US-08-171-718-104	Sequence 104, App

ALIGNMENTS

```
RESULT 1
US-08-247-901C-1/c
; Sequence 1, Application US/08247901C
; Patent No. 5750384
; GENERAL INFORMATION:
;   APPLICANT: Jacobs et al
;   TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Amster, Rothstein & Ebenstein
;     STREET: 90 Park Avenue
;     CITY: New York
;     STATE: New York
;     COUNTRY: U.S.A.
;     ZIP: 10016
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
;     COMPUTER: IBM PC Compatible
;     OPERATING SYSTEM: MS-DOS
;     SOFTWARE: Word Processor (ASCII)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/247,901C
;     FILING DATE: May 23, 1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/057,531
;     FILING DATE: April 29, 1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Bogosian, Elizabeth A
;     REGISTRATION NUMBER: 39,911
;     REFERENCE/DOCKET NUMBER: 96700/273
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 697-5995
;     TELEFAX: (212) 286-0854 or 286-0082
;     TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 50341
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE:
;     DESCRIPTION: L5 shuttle phasmid sequence
;   HYPOTHETICAL: No
;   ANTI-SENSE:
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
;     ORGANISM: L5 mycobacteriophage
;     STRAIN:
;     INDIVIDUAL ISOLATE:
```

```
;       DEVELOPMENTAL STAGE:
;       HAPLOTYPE:
;       TISSUE TYPE:
;       CELL TYPE:
;       CELL LINE:
;       ORGANELLE:
;     IMMEDIATE SOURCE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:  No. 5750384e
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1


  Query Match             76.0%;  Score 15.2;  DB 2;  Length 50341;
  Best Local Similarity   85.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGCCGACATCATCGCTTCCC 20
           |||||| |||||| |||| |
Db    8578 GGCCGAGATCATCTCTTCTC 8559


RESULT   2
US-09-075-904-1/c
; Sequence 1, Application US/09075904
; Patent No. 5994137
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, et al.
;    TITLE OF INVENTION:  L5 SHUTTLE PHASMIDS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amster, Rothstein & Ebenstein
;      STREET:  90 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,904
;      FILING DATE:  May 11, 1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/247,901
;      FILING DATE:  May 23, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Elizabeth A
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/475
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50341
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  L5 shuttle phasmid sequence
;    HYPOTHETICAL:  No
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  L5 mycobacteriophage
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:  No. 5994137e
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN   Q ID NO:
US-09-075-904-1


  Query Match             76.0%;  Score 15.2;  DB 4;  Length 50341;
  Best Local Similarity   85.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGCCGACATCATCGCTTCCC 20
           |||||| |||||| |||| |
Db    8578 GGCCGAGATCATCTCTTCTC 8559


RESULT   3
US-08-530-569B-20/c
; Sequence 20, Application US/08530569B
; Patent No. 5939526
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, Beatrice
;    APPLICANT:  van den Eynde, Benoit
;    APPLICANT:  Schrier, Peter
;    APPLICANT:  Brouwenstijn, Nathalie
;    APPLICANT:  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated RAGE-1 Derived Peptides Which
;    TITLE OF INVENTION:  Complex with HLA-B7 Molecules and Uses Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  Federal Reserve Plaza, 600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/530,569B
;       FILING DATE:
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gates, Edward R.
;       REGISTRATION NUMBER:  31,616
;       REFERENCE/DOCKET NUMBER:  L0461/7002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-720-3500
;       TELEFAX:  617-720-2441
;   INFORMATION FOR SEQ ID NO:  20:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  189 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..189
US-08-530-569B-20


  Query Match             71.0%;  Score 14.2;  DB 4;  Length 189;
  Best Local Similarity   84.2%;  Pred. No. 73;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGCCGACATCATCGCTTCC 19
           ||||  |||| ||||||||
Db      93 GGCCTTCATCCTCGCTTCC 75


RESULT   4
US-08-867-030B-2
; Sequence 2, Application US/08867030B
; Patent No. 5948900
;  GENERAL INFORMATION:
;    APPLICANT:  Yother et al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae
;    TITLE OF INVENTION:  Capsular Polysaccharide Genes and Flanking Regions
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Benjamin Aaron Adler, Ph.D., J.D.
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch floppy
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,030B
;      FILING DATE:  June 2, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/243,546
;      FILING DATE:  May 16, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benjamin Aaron Adler, Ph.D., J.D.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5923
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 777-2321
;      TELEFAX:  (713) 777-6908
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  261 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single-stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:
US-08-867-030B-2


  Query Match             71.0%;  Score 14.2;  DB 4;  Length 261;
  Best Local Similarity   84.2%;  Pred. No. 76;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           ||||| |||| |||||| |
Db     115 GCCGATATCACCGCTTCGC 133


RESULT   5
PCT-US95-06119-2
; Sequence 2, Application PC/TUS9506119
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  STREPTOCOCCUS PNEUMONIAE CAPSULAR
;    TITLE OF INVENTION:  POLYSACCHARIDE GENES AND FLANKING REGIONS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version
;      SOFTWARE:  #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06119
;      FILING DATE:  CONCURRENTLY HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,546
;      FILING DATE:  16-MAY-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  AMCY018P--
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  261 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
PCT-US95-06119-2
```

```
  Query Match             71.0%;  Score 14.2;  DB 6;  Length 261;
  Best Local Similarity   84.2%;  Pred. No. 76;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           ||||| |||| |||||| |
Db     115 GCCGATATCACCGCTTCGC 133


RESULT   6
US-08-726-136-21
; Sequence 21, Application US/08726136
; Patent No. 5811286
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERT D. FALLON
;    APPLICANT:  MARK S. PAYNE
;    APPLICANT:  MARK J. NELSON
;    TITLE OF INVENTION:  NUCLEIC ACID FRAGMENTS ENCODING
;    TITLE OF INVENTION:  STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
;    TITLE OF INVENTION:  RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
;    TITLE OF INVENTION:  THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,136
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/004914
;      FILING DATE:  OCTOBER 6, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA A.
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      STRAIN:  P14K
US-08-726-136-21


  Query Match             71.0%;  Score 14.2;  DB 2;  Length 384;
  Best Local Similarity   84.2%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           ||||| |||  ||||||||
Db     108 GCCGAAATCGACGCTTCCC 126


RESULT   7
US-08-581-528A-5/c
; Sequence 5, Application US/08581528A
; Patent No. 5986058
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, Se-Jin
;    APPLICANT:  Huynh, Thanh
;    TITLE OF INVENTION:  GROWTH DIFFERENTIATION FACTOR-7
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/581,528A
;      FILING DATE:  03-Sept-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/089,670
;      FILING DATE:  09-JUL-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lisa A. Haile, Ph.D.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  07265/081001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  519 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  GDF-7
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  34..516
US-08-581-528A-5


  Query Match             71.0%;  Score 14.2;  DB 4;  Length 519;
  Best Local Similarity   84.2%;  Pred. No. 83;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           |||| || || ||||||||
Db     126 GCCGCCACCACCGCTTCCC 108


RESULT   8
PCT-US94-07799-5/c
; Sequence 5, Application PC/TUS9407799
;  GENERAL INFORMATION:
;    APPLICANT:  THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
;    TITLE OF INVENTION:  GROWTH DIFFERENTIATION FACTOR-7
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
```

```
;       ZIP:  90067
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/07799
;       FILING DATE:  08-JUL-1994
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  TUMARKIN, LISA A., PH.D.
;       REGISTRATION NUMBER:  P-38,347
;       REFERENCE/DOCKET NUMBER:  FD-2348
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 455-5100
;       TELEFAX:  (619) 455-5110
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  519 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  GDF-7
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  34..516
PCT-US94-07799-5


  Query Match             71.0%;  Score 14.2;  DB 6;  Length 519;
  Best Local Similarity   84.2%;  Pred. No. 83;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           |||| || || ||||||||
Db     126 GCCGCCACCACCGCTTCCC 108


RESULT  9
US-08-530-569B-18/c
; Sequence 18, Application US/08530569B
; Patent No. 5939526
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, Beatrice
;    APPLICANT:  van den Eynde, Benoit
;    APPLICANT:  Schrier, Peter
;    APPLICANT:  Brouwenstijn, Nathalie
;    APPLICANT:  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated RAGE-1 Derived Peptides Which
;    TITLE OF INVENTION:  Complex with HLA-B7 Molecules and Uses Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  Federal Reserve Plaza, 600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,569B
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gates, Edward R.
;      REGISTRATION NUMBER:  31,616
;      REFERENCE/DOCKET NUMBER:  L0461/7002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  564 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..564
US-08-530-569B-18


  Query Match             71.0%;  Score 14.2;  DB 4;  Length 564;
  Best Local Similarity   84.2%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGCCGACATCATCGCTTCC 19
           ||||  |||| ||||||||
Db     193 GGCCTTCATCCTCGCTTCC 175


RESULT  10
US-08-362-670B-29/c
; Sequence 29, Application US/08362670B
; Patent No. 5658882
;  GENERAL INFORMATION:
;    APPLICANT:  Celeste, Anthony J.
;    APPLICANT:  Wozney, John
;    APPLICANT:  Rosen, Vicki A.
;    APPLICANT:  Wolfman, Neil
;    APPLICANT:  Thomsen, Gerald H.
;    APPLICANT:  Melton, Douglas A.
;    TITLE OF INVENTION:  TENDON-INDUCING COMPOSITIONS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,670B
;      FILING DATE:  December 22, 1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5202-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 498-8260
;      TELEFAX:  617 876-5851
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1203 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
```

```
;      CLONE:  murine MV1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..721
US-08-362-670B-29


  Query Match             71.0%;  Score 14.2;  DB 1;  Length 1203;
  Best Local Similarity   84.2%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GCCGACATCATCGCTTCCC 20
          |||| || || ||||||||
Db    331 GCCGCCACCACCGCTTCCC 313


RESULT  11
US-08-333-576C-29/c
; Sequence 29, Application US/08333576C
; Patent No. 6027919
;  GENERAL INFORMATION:
;    APPLICANT:  Celeste, Anthony J.
;    APPLICANT:  Wozney, John
;    APPLICANT:  Rosen, Vicki A.
;    APPLICANT:  Wolfman, Neil
;    APPLICANT:  Thomsen, Gerald H.
;    APPLICANT:  Melton, Douglas A.
;    TITLE OF INVENTION:  TENDON-INDUCING COMPOSITIONS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/333,576C
;      FILING DATE:  No. 6027919ember 2, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5202-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 498-8260
;      TELEFAX:  617 876-5851
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1203 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  murine MV1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..721
US-08-333-576C-29


  Query Match             71.0%;  Score 14.2;  DB 5;  Length 1203;
  Best Local Similarity   84.2%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GCCGACATCATCGCTTCCC 20
          |||| || || ||||||||
Db    331 GCCGCCACCACCGCTTCCC 313


RESULT  12
PCT-US94-14030A-29/c
; Sequence 29, Application PC/TUS9414030A
;  GENERAL INFORMATION:
;    APPLICANT:  GENETICS INSTITUTE, INC.
;    APPLICANT:  PRESIDENT AND FELLOWS OF HARVARD COLLEGE
;    TITLE OF INVENTION:  TENDON-INDUCING COMPOSITIONS
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/14030A
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/164,103
;      FILING DATE:  07-DEC-1993
;      APPLICATION NUMBER:  US 08/217,780
;      FILING DATE:  25-MAR-1994
;      APPLICATION NUMBER:  US 08/333,576
;      FILING DATE:  02-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5202D-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 498-8260
;      TELEFAX:  617 876-5851
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1203 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  murine MV1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..721
PCT-US94-14030A-29


  Query Match             71.0%;  Score 14.2;  DB 6;  Length 1203;
  Best Local Similarity   84.2%;  Pred. No. 92;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 GCCGACATCATCGCTTCCC 20
          |||| || || ||||||||
Db    331 GCCGCCACCACCGCTTCCC 313


RESULT  13
US-08-530-569B-13/c
; Sequence 13, Application US/08530569B
; Patent No. 5939526
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, Beatrice
;    APPLICANT:  van den Eynde, Benoit
;    APPLICANT:  Schrier, Peter
```

```
;    APPLICANT:  Brouwenstijn, Nathalie
;    APPLICANT:  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated RAGE-1 Derived Peptides Which
;    TITLE OF INVENTION:  Complex with HLA-B7 Molecules and Uses Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  Federal Reserve Plaza, 600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,569B
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gates, Edward R.
;      REGISTRATION NUMBER:  31,616
;      REFERENCE/DOCKET NUMBER:  L0461/7002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2051 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-530-569B-13


  Query Match             71.0%;  Score 14.2;  DB 4;  Length 2051;
  Best Local Similarity   84.2%;  Pred. No. 99;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGCCGACATCATCGCTTCC 19
           ||||  |||| ||||||||
Db     461 GGCCTTCATCCTCGCTTCC 443


RESULT  14
US-08-883-534-4/c
; Sequence 4, Application US/08883534
; Patent No. 5846777
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  TWO NEW WD-40 PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,534
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0332 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2369 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  NEUTGMT01
;      CLONE:  1221143
US-08-883-534-4


  Query Match             71.0%;  Score 14.2;  DB 3;  Length 2369;
  Best Local Similarity   84.2%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 GCCGACATCATCGCTTCCC 20
           || |  |||||||||||||
Db     628 GCAGTTATCATCGCTTCCC 610


RESULT  15
US-09-204-764-4/c
; Sequence 4, Application US/09204764
; Patent No. 6025464
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  TWO NEW WD-40 PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/204,764
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/883,534
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0332 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: NEUTGMT01
; CLONE: 1221143
US-09-204-764-4

Query Match 71.0%; Score 14.2; DB 5; Length 2369;
Best Local Similarity 84.2%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 GCCGACATCATCGCTTCCC 20
           || |  ||||||||||||
Db     628 GCAGTTATCATCGCTTCCC 610
```

Search completed: June 22, 2000, 18:04:01
Job time: 9708 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:30 ; Search time 2113.77 Seconds
(without alignments)
38.351 Million cell updates/sec

Title: US-09-362-485-22
Perfect score: 20
Sequence: 1 GGCCGACATCATCGCTTCCC 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	16.8	84.0	619	82	FR0013267	AL004517 F.rubripe
	2	16.4	82.0	286	41	AI059899	AI059899 UI-R-C1-1
	3	16.4	82.0	324	36	C74164	C74164 C74164 Rice
	4	15.8	79.0	257	69	AW115447	AW115447 rs67f12.y
	5	15.8	79.0	300	33	AA442579	AA442579 zv57f03.r
	6	15.8	79.0	360	69	AW114600	AW114600 rs52d12.y
	7	15.8	79.0	368	70	AW145010	AW145010 rs44e11.y
	8	15.8	79.0	375	36	C71050	C71050 C71050 Yuji
	9	15.8	79.0	390	69	AW114843	AW114843 rs60c03.y
c	10	15.8	79.0	406	41	AI047265	AI047265 uh63d12.r
	11	15.8	79.0	424	25	N94598	N94598 zb79b08.s1
c	12	15.8	79.0	430	33	AA420776	AA420776 nc63e10.s
	13	15.8	79.0	430	64	AW052338	AW052338 rs34c11.y
	14	15.8	79.0	438	26	W85882	W85882 zh56e04.s1
	15	15.8	79.0	451	43	AI203699	AI203699 qf54c03.x
	16	15.8	79.0	454	40	AA932824	AA932824 oo64a06.s
c	17	15.8	79.0	455	61	AI813122	AI813122 23C10 Pin
	18	15.8	79.0	457	43	AI224497	AI224497 qi34f03.x
	19	15.8	79.0	458	69	AW114734	AW114734 rs54g04.y
	20	15.8	79.0	461	64	AW097006	AW097006 rs31e08.y
c	21	15.8	79.0	464	30	AA255975	AA255975 zs29d06.r
	22	15.8	79.0	465	42	AI141816	AI141816 qa56f12.s
c	23	15.8	79.0	468	63	AI986825	AI986825 rs17e09.y
c	24	15.8	79.0	473	33	AA420444	AA420444 nc60b10.s
c	25	15.8	79.0	477	79	AW269345	AW269345 xs47e12.x
	26	15.8	79.0	491	36	AA643378	AA643378 nr59f03.s
	27	15.8	79.0	497	33	AA420799	AA420799 nc63e10.r
	28	15.8	79.0	541	33	AA420443	AA420443 nc60b10.r
	29	15.8	79.0	575	51	AI742159	AI742159 wg49f07.x
	30	15.8	79.0	577	60	AI799690	AI799690 tm80a09.x
	31	15.8	79.0	589	41	AI051243	AI051243 oy90d05.x
c	32	15.8	79.0	591	103	AQ257746	AQ257746 nbxb0018N
	33	15.8	79.0	605	51	AI733568	AI733568 ou47b07.x
c	34	15.8	79.0	615	37	AB009159	AB009159 AB009159
	35	15.8	79.0	616	44	AI264192	AI264192 qi20g11.x
	36	15.8	79.0	686	41	AI018507	AI018507 ou47b07.x
	37	15.8	79.0	991	83	AF076009	AF076009 Salmonell
c	38	15.4	77.0	336	49	AI639905	AI639905 vr72d04.x
	39	15.4	77.0	401	107	AQ445576	AQ445576 GSSTc0031
c	40	15.4	77.0	412	37	AA675479	AA675479 vr72d04.s
	41	15.4	77.0	444	91	AQ864687	AQ864687 nbeb0023P
c	42	15.4	77.0	525	74	AW174891	AW174891 fe06h06.y
	43	15.4	77.0	543	26	W44606	W44606 zc29c04.r1
c	44	15.4	77.0	627	49	AI649511	AI649511 603005E08
	45	15.2	76.0	195	72	AV297715	AV297715 AV297715

ALIGNMENTS

RESULT 1
FR0013267/c
LOCUS FR0013267 619 bp DNA GSS 18-SEP-1997
DEFINITION F.rubripes GSS sequence, clone 113O17bA2, genomic survey sequence.
ACCESSION AL004517
VERSION AL004517.1 GI:2450087
KEYWORDS GSS; genome survey sequence.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Fugu.
REFERENCE 1 (bases 1 to 619)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES Location/Qualifiers
source 1. .619
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 113O17"
/clone="113O17bA2"
BASE COUNT 159 a 162 c 137 g 120 t 41 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 82; Length 619;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCCC 20
||||||||||||||||| | |
Db 260 GGCCGACATCATCGCTGCTC 241

RESULT 2
AI059899
LOCUS AI059899 286 bp mRNA EST 05-JUL-1999
DEFINITION UI-R-C1-1a-g-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
UI-R-C1-1a-g-11-0-UI 3', mRNA sequence.
ACCESSION AI059899
VERSION AI059899.1 GI:3333676
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 286)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900209.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA.
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult Ovary library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1775752
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .286
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-1a-g-11-0-UI"
/clone_lib="UI-R-C1"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.

The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 61 a 72 c 63 g 90 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 41; Length 286;
Best Local Similarity 94.4%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCCGACATCATCGCTTCC 19
||||||||||||| |||||
Db 250 GCCGACATCATCCCTTCC 267

RESULT 3
C74164
LOCUS C74164 324 bp mRNA EST 29-SEP-1997
DEFINITION C74164 Rice panicle shorter than 3cm Oryza sativa cDNA clone E30539_1A, mRNA sequence.
ACCESSION C74164
VERSION C74164.1 GI:2442393
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 324)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle
JOURNAL Unpublished (1997)
COMMENT On May 8, 1995 this sequence version replaced gi:801440.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
FEATURES Location/Qualifiers
source 1. .324
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E30539_1A"
/clone_lib="Rice panicle shorter than 3cm"
/dev_stage="shorter than 3cm"
/note="Organ: panicle"
BASE COUNT 49 a 110 c 100 g 56 t 9 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 36; Length 324;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CCGACATCATCGCTTCCC 20
||||| ||||||||||||
Db 43 CCGACTTCATCGCTTCCC 60

RESULT 4
AW115447
LOCUS AW115447 257 bp mRNA EST 20-OCT-1999
DEFINITION rs67f12.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5', mRNA sequence.
ACCESSION AW115447
VERSION AW115447.1 GI:6081785
KEYWORDS EST.
SOURCE Pristionchus pacificus.
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Diplogasterida; Diplogasterina; Diplogasteroidea; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 257)
AUTHORS McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT On Mar 16, 1998 this sequence version replaced gi:2961776.
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
Washington University Genome Sequencing Center
Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for information about this clone.
Putative full length read
The vector to vector length is 278
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .257
/organism="Pristionchus pacificus"
/strain="PS 312"
/db_xref="taxon:54126"
/clone_lib="Sommer Pristionchus"
/sex="predominantly hermaphroditic"
/dev_stage="mixed stages (embryo to adult)"
/lab_host="not applicable (host cell line)"
/note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5' EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a XhoI - oligo(dT) primer. Double-stranded cDNA was ligated to EcoRI adaptors digested with XhoI and cloned into XhoI and EcoRI sites. Primary complexity of the library was 10 in the 7th. The library went through one round of amplification."
BASE COUNT 70 a 60 c 75 g 51 t 1 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 69; Length 257;
Best Local Similarity 89.5%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCC 19
|||||| ||||||||| ||
Db 31 GGCCGAAATCATCGCTCCC 49

```
RESULT   5
AA442579
LOCUS       AA442579      300 bp    mRNA            EST       02-JUN-1997
DEFINITION  zv57f03.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757757
            5', mRNA sequence.
ACCESSION   AA442579
VERSION     AA442579.1  GI:2154457
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404537.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28ml3 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Homo sapiens"
                     /db_xref="GDB:5978371"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:757757"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       83 a     87 c     75 g     55 t
ORIGIN


  Query Match           79.0%;  Score 15.8;  DB 33;  Length 300;
  Best Local Similarity  89.5%;  Pred. No. 4.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 GCCGACATCATCGCTTCCC 20
          ||||||||||||| || ||
Db    242 GCCGACATCATCGATTTCC 260


RESULT   6
AW114600
LOCUS       AW114600      360 bp    mRNA            EST       20-OCT-1999
DEFINITION  rs52d12.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5'
            similar to WP:R151.3 CE00744 RIBOSOMAL PROTEIN ML16 ;, mRNA
            sequence.
ACCESSION   AW114600
VERSION     AW114600.1  GI:6080875
KEYWORDS    EST.
SOURCE      Pristionchus pacificus.
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria;
            Diplogasterida; Diplogasterina; Diplogasteroidea;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M.,
            Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S.,
            Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247509.
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
            Washington University Genome Sequencing Center
             Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for
            information about this clone.
            Putative full length read
            The vector to vector length is 361
            Seq primer: -40RP from Gibco
            High quality sequence stop: 261.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Pristionchus pacificus"
                     /strain="PS 312"
                     /db_xref="taxon:54126"
                     /clone_lib="Sommer Pristionchus"
                     /sex="predominantly hermaphroditic"
                     /dev_stage="mixed stages (embryo to adult)"
                     /lab_host="not applicable (host cell line)"
                     /note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5'
                     EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a
                     XhoI - oligo(dT) primer. Double stranded cDNA was ligated
                     to EcoRI adaptors digested with XhoI and cloned into XhoI
                     and EcoRI sites. Primary complexity of the library was 10
                     in the 7th. The library went through one round of
                     amplification."
BASE COUNT       90 a     100 c     90 g     79 t      1 others
ORIGIN


  Query Match           79.0%;  Score 15.8;  DB 69;  Length 360;
  Best Local Similarity  89.5%;  Pred. No. 5e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCC 19
          ||||||||||| |||| |||
Db    117 GGCCGACATCTTCGCATCC 135


RESULT   7
AW145010
LOCUS       AW145010      368 bp    mRNA            EST       01-NOV-1999
DEFINITION  rs44e11.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5'
            similar to WP:R151.3 CE00744 RIBOSOMAL PROTEIN ML16 ;contains
            element MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AW145010
VERSION     AW145010.1  GI:6166746
KEYWORDS    EST.
SOURCE      Pristionchus pacificus.
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria;
            Diplogasterida; Diplogasterina; Diplogasteroidea;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,
```

Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5406880.
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
Washington University Genome Sequencing Center
Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for information about this clone.
Seq primer: -40RP from Gibco
High quality sequence stop: 173.
FEATURES Location/Qualifiers
source 1. .368
/organism="Pristionchus pacificus"
/strain="PS 312"
/db_xref="taxon:54126"
/clone_lib="Sommer Pristio chus"
/sex="predominantly hermap..roditic"
/dev_stage="mixed stages (embryo to adult)"
/lab_host="not applicable (host cell line)"
/note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5' EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a XhoI - oligo(dT) primer. Double-stranded cDNA was ligated to EcoRI adaptors digested with XhoI and cloned into XhoI and EcoRI sites. Primary complexity of the library was 10 in the 7th. The library went through one round of amplification."
BASE COUNT 88 a 112 c 95 g 72 t 1 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 70; Length 368;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCC 19
|||||||||| |||| |||
Db 206 GGCCGACATCTTCGCATCC 224

RESULT 8
C71050
LOCUS C71050 375 bp mRNA EST 29-SEP-1997
DEFINITION C71050 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk439b9 5', mRNA sequence.
ACCESSION C71050
VERSION C71050.1 GI:2441575
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 375)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Apr 18, 1995 this sequence version replaced gi:775516.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES Locat on/Qualifiers
source 1. .375
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk439b9"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 85 a 94 c 112 g 82 t 2 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 36; Length 375;
Best Local Similarity .5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCGACATCATCGCTTCCC 20
||||| |||| ||||||||
Db 245 GCCGAGATCAACGCTTCCC 263

RESULT 9
AW114843
LOCUS AW114843 390 bp mRNA EST 20-OCT-1999
DEFINITION rs60c03.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5' similar to WP:R151.3 CE00744 RIBOSOMAL PROTEIN ML16 ;, mRNA sequence.
ACCESSION AW114843
VERSION AW114843.1 GI:6081181
KEYWORDS EST.
SOURCE Pristionchus pacificus.
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria; Diplogasterida; Diplogasterina; Diplogasteroidea; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 390)
AUTHORS McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT On Feb 18, 1999 this sequence version replaced gi:4299406.
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
Washington University Genome Sequencing Center
Contact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for information about this clone.
Putative full length read
The vector to vector length is 428
Seq primer: -40RP from Gibco
High quality sequence stop: 372.
FEATURES Location/Qualifiers
source 1. .390
/organism="Pristionchus pacificus"
/strain="PS 312"
/db_xref="taxon:54126"
/clone_lib="Sommer Pristionchus"
/sex="predominantly hermaphroditic"
/dev_stage="mixed stages (embryo to adult)"
/lab_host="not applicable (host cell line)"

/note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5' EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a XhoI - oligo(dT) primer. Double-stranded cDNA was ligated to EcoRI adaptors digested with XhoI and cloned into XhoI and EcoRI sites. Primary complexity of the library was 10 in the 7th. The library went through one round of amplification."
BASE COUNT 104 a 126 c 91 g 69 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 60; Length 390;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCC 19
||||||||||| |||| |||
Db 197 GGCCGACATCTTCGCATCC 215

RESULT 10
AI047265/c
LOCUS AI047265 406 bp mRNA EST 08-JUL-1998
DEFINITION uh63d12.r1 Soares_embryonic_stem_cell_NMES Mus musculus cDNA clone IMAGE:1750103 5' similar to TR:O04892 O04892 CYTOCHROME P450 LIKE_TBP ;, mRNA sequence.
ACCESSION AI047265
VERSION AI047265.1 GI:3295552
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 406)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 8, 1995 this sequence version replaced gi:800273.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:961915
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 236.
FEATURES Location/Qualifiers
source 1. .406
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1750103"
/clone_lib="Soares_embryonic_stem_cell_NMES"
/cell_type="embryonic stem cell"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 81 a 121 c 137 g 64 t 3 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 41; Length 406;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCCGACATCATCGCTTCCC 20
||| |||||||||||| |
Db 301 GCCCACATCATCGCTTCTC 283

RESULT 11
N94598
LOCUS N94598 424 bp mRNA EST 20-AUG-1996
DEFINITION zb79b08.s1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:309783 3', mRNA sequence.
ACCESSION N94598
VERSION N94598.1 GI:1266907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1461 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 255.
FEATURES Location/Qualifiers
source 1. .424
/organism="Homo sapiens"
/db_xref="GDB:1253197"
/db_xref="taxon:9606"
/clone="IMAGE:309783"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 81 a 118 c 113 g 110 t 2 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 25; Length 424;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 GGCCGACATCATCGCTTCC 19
          ||| |||||||| ||||||
Db    400 GGCAGACATCATGGCTTCC 418


RESULT  12
AA420776/c
LOCUS       AA420776      430 bp    mRNA            EST       16-OCT-1997
DEFINITION  nc63e10.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:745962
            similar to TR:G662338 G662338 SIMILAR TO C. ELEGANS HYPOTHETICAL
            PROTEIN C34E10.2. ;, mRNA sequence.
ACCESSION   AA420776
VERSION     AA420776.1  GI:2094638
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence versio. replaced gi:140079.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 889   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="GDB:5923292"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:745962"
                     /clone_lib="NCI_CGAP_Pr1"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected, histologically normal
                     prostate epithelial cells. Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by David Krizman."
BASE COUNT       )9 a    135 c     99 g     97 t
ORIGIN


  Query Match           79.0%;  Score 15.8;  DB 33;  Length 430;
  Best Local Similarity  89.5%;  Pred. No. 5.3e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCC 19
          ||| |||||||| ||||||
Db    352 GGCAGACATCATGGCTTCC 334


RESULT  13
AW052338
LOCUS       AW052338      430 bp    mRNA            EST       20-SEP-1999
DEFINITION  rs34c11.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5'
            similar to WP:R151.3 CE00744 RIBOSOMAL PROTEIN ML16 ;, mRNA
            sequence.
ACCESSION   AW052338
VERSION     AW052338.1  GI:5914697
KEYWORDS    EST.
SOURCE      Pristionchus pacificus.
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria;
            Diplogasterida; Diplogasterina; Diplogasteroidea;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   McCarter,J., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M.,
            Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S.,
            Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135803.
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Ralf Sommer DNA Sequencing by:
            W shington University Genome Sequencing Center
             ontact Dr. Ralf Sommer (ralf.sommer@tuebingen.mpg.de) for
            i formation about this clone.
            Putative full length read
            The vector to vector length is 431
            Seq primer: -40RP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Pristionchus pacificus"
                     /strain="PS 312"
                     /db_xref="taxon:54126"
                     /clone_lib="Sommer Pristionchus"
                     /sex="predominantly hermaphroditic"
                     /dev_stage="mixed stages (embryo to adult)"
                     /lab_host="not applicable (host cell line)"
                     /note="Vector: Uni-ZAP XR Vector (Stratagene); Site_1: 5'
                     EcoRI; Site_2: 3' XhoI; 1st strand cDNA was primed with a
                     XhoI - oligo(dT) primer. Double-stranded cDNA was ligated
                     to EcoRI adaptors digested with XhoI and cloned into XhoI
                     and EcoRI sites. Primary complexity of the library was 10
                     in the 7th. The library went through one round of
                     amplification."
BASE COUNT       113 a    132 c     98 g     87 t
ORIGIN


  Query Match           79.0%;  Score 15.8;  DB 64;  Length 430;
  Best Local Similarity  89.5%;  Pred. No. 5.3e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCCGACATCATCGCTTCC 19
          |||||||||| |||| |||
Db    206 GGCCGACATCTTCGCATCC 224


RESULT  14
W85882
LOCUS       W85882        438 bp    mRNA            EST       02-FEB-1997
DEFINITION  zh56e04.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
```

clone IMAGE:416094 3', mRNA sequence.
ACCESSION W85882
VERSION W85882.1 GI:1398311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On May 5, 1995 this sequence version replaced gi:797723.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501 St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 496 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 336.
FEATURES Location/Qualifiers
source 1. .438
/organism="Homo sapiens"
/db_xref="GDB:1324564"
/db_xref="taxon:9606"
/clone="IMAGE:416094"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 78 a 127 c 120 g 113 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 26; Length 438;
Best Local Similarity 89.5%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCC 19
||| |||||||| ||||||
Db 378 GGCAGACATCATGGCTTCC 396

RESULT 15
AI203699
LOCUS AI203699 451 bp mRNA EST 14-OCT-1998
DEFINITION qf54c03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1753828
3', mRNA sequence.
ACCESSION AI203699
VERSION AI203699.1 GI:3756305
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 451)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151987.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 425.
FEATURES Location/Qualifiers
source 1. .451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1753828"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 85 a 128 c 120 g 118 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 43; Length 451;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCCGACATCATCGCTTCC 19
||| |||||||| ||||||
Db 401 GGCAGACATCATGGCTTCC 419

Search completed: June 22, 2000, 17:43:34
Job time: 20084 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:13:08 ; Search time 175.2 Seconds
(without alignments)
25.705 Million cell updates/sec

Title: US-09-362-485-21
Perfect score: 18
Sequence: 1 CGCGGCCGACATCATCGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49522	Mycobacterium sp.
2		18	100.0	18	1	V49620	AlaDH derived olig
3	c	18	100.0	1245	1	V49626	Mycobacterium tube
4	c	18	100.0	1260	1	V49510	Mycobacterium sp.
5	c	18	100.0	1260	1	V49625	Mycobacterium tube
6		16	88.9	1155	1	T91517	Mycobacterium tube
7		16	88.9	1155	1	T91453	Mycobacterium tube
8		16	88.9	1155	1	V44350	Mycobacterium tube
9		16	88.9	1155	1	V64458	M. tuberculosis im
10		15.4	85.6	12019	1	X18867	Alcaligenes sp. Po
11		15.4	85.6	13144	1	Q13288	P.denitrificans ge
12		15	83.3	20	1	V49523	Mycobacterium sp.
13		15	83.3	20	1	V49621	AlaDH derived olig
14		15	83.3	2664	1	V59465	Nucleotide sequenc
15	c	15	83.3	6317	1	Q55756	Enterobacter cloac
16	c	14.8	82.2	1185	1	V05039	Alcaligenes eutrop
17	c	14.8	82.2	1348	1	Q67407	MyrB gene DNA, enc
18		14.8	82.2	3287	1	X26587	Nucleotide sequenc
19	c	14.8	82.2	20387	1	V62159	HSV-2 strain SB5 C
20	c	14.8	82.2	26338	1	V62134	HSV-2 strain SB5 C
21		14.8	82.2	53789	1	V21187	Amycolatopsis medi
22	c	14.8	82.2	117213	1	V62176	HSV-2 strain SB5 C
23		14.4	80.0	2715	1	V64624	P. putida cis/tran
24		14.4	80.0	6575	1	Q84655	Human neuronal cal
25		14.4	80.0	6725	1	Q87834	Human neuronal cal
26	c	14.4	80.0	29879	1	Q46806	eryA region of S.
27	c	14	77.8	2098	1	T12651	pAC22 cDNA encodin
28	c	14	77.8	2110	1	T12652	pAC92 cDNA encodin
29		14	77.8	4131	1	T95860	Rat kidney cell ca
30		14	77.8	4131	1	V26965	Rat kidney calcium
31		14	77.8	4131	1	V82486	Rat parathyroid ca
32		13.8	76.7	22	1	T94437	Primer SEQ ID NO:1
33		13.8	76.7	53	1	N92908	Synthetic double s
34		13.8	76.7	84	1	T06317	Thaumatin II synth
35	c	13.8	76.7	414	1	X41342	Human secreted pro
36	c	13.8	76.7	486	1	X34091	Mycobacterium spec
37		13.8	76.7	624	1	T06312	Thaumatin II synth
38		13.8	76.7	687	1	N90625	Sequence encoding
39		13.8	76.7	687	1	Q04276	PBlmn, a HIV fusio
40		13.8	76.7	811	1	V44446	Mycobacterium tube
41		13.8	76.7	811	1	V64555	M. tuberculosis im
42		13.8	76.7	1026	1	T79497	Meripilus giganteu
43		13.8	76.7	1323	1	V54704	Mycobacterium tube
44		13.8	76.7	1323	1	V54705	Mycobacterium tube
45		13.8	76.7	1323	1	V54706	Mycobacterium tube

ALIGNMENTS

RESULT 1
V49522
ID V49522 standard; DNA; 18 BP.
AC V49522;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-R4.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 3 A; 8 C; 5 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
||||||||||||||||||
Db 1 CGCGGCCGACATCATCGC 18

RESULT 2
V49620
ID V49620 standard; DNA; 18 BP.
AC V49620;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-R4.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 3 A; 8 C; 5 G; 2 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
||||||||||||||||||
Db 1 CGCGGCCGACATCATCGC 18

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
||||||||||||||||||
Db 1245 CGCGGCCGACATCATCGC 1228

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGGCCGACATCATCGC 18
||||||||||||||||||
Db 1260 CGCGGCCGACATCATCGC 1243

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match             100.0%;  Score 18;  DB 1;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 2.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATCGC 18
          ||||||||||||||||||
Db   1260 CGCGGCCGACATCATCGC 1243


RESULT   6
T91517
ID   T91517 standard; DNA; 1155 BP.
AC   T91517;
DT   08-JAN-1998  (first entry)
DE   Mycobacterium tuberculosis antigen TbRa28 encoding DNA.
KW   Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW   skin testing; M.tuberculosis; ss.
OS   Mycobacterium tuberculosis.
FH   Key             Location/Qualifiers
FT   CDS             2. .928
FT                   /*tag= a
FT                   /product= Antigen_TbRa28
PN   WO9709428-A2.
PD   13-MAR-1997.
PF   30-AUG-1996; U14674.
PR   12-JUL-1996; US-680574.
PR   01-SEP-1995; US-523436.
PR   22-SEP-1995; US-533634.
PR   22-MAR-1996; US-620874.
PR   05-JUN-1996; US-659683.
PA   (CORI-) CORIXA CORP.
PI   Campos-neto A,  Dillon DC,  Houghton R,  Reed SG,  Skeiky YAW;
PI   Twardzik DR, Vedvick TH;
DR   WPI; 97-192903/17.
DR   P-PSDB; W32430.
PT   New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT   useful in vaccines for prevention or treatment of tuberculosis, also
PT   for diagnosis
PS   Claim 11; Page 58; 168pp; English.
CC   A new immunogenic polypeptide has been developed comprising an
CC   immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC   its variant differing only in conservative substitutions and/or
CC   modifications). The present sequence encodes a M.tuberculosis antigen,
CC   TbRa28, which can be used in part of a pharmaceutical composition
CC   (claimed). The immunogenic protein, and fusion proteins containing one
CC   or more of the proteins or one of the proteins plus ESAT-6, are useful
CC   in vaccines, preferably when formulated with a non-specific adjuvant,
CC   to induce an immune response against M.tuberculosis (for treatment or
CC   prevention).
SQ   Sequence   1155 BP;    203 A;    377 C;    394 G;    181 T;


  Query Match             88.9%;  Score 16;  DB 1;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATC 16
          ||||||||||||||||
Db    806 CGCGGCCGACATCATC 821


RESULT   7
T91453
ID   T91453 standard; DNA; 1155 BP.
AC   T91453;
DT   13-JAN-1998  (first entry)
DE   Mycobacterium tuberculosis antigen TbRa28 encoding DNA.
KW   Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW   skin testing; M.tuberculosis; ss.
OS   Mycobacterium tuberculosis.
FH   Key             Location/Qualifiers
FT   CDS             2. .928
FT                   /*tag= a
FT                   /product= Antigen_TbRa28
PN   WO9709429-A2.
PD   13-MAR-1997.
PF   30-AUG-1996; U14675.
PR   12-JUL-1996; US-680573.
PR   01-SEP-1995; US-523435.
PR   22-SEP-1995; US-532136.
PR   22-MAR-1996; US-620280.
PR   05-JUN-1996; US-658800.
PA   (CORI-) CORIXA CORP.
PI   Campos-neto A,  Dillon DC,  Houghton R,  Reed SG,  Skeiky YAW;
PI   Twardzik DR, Vedvick TH;
DR   WPI; 97-192904/17.
DR   P-PSDB; W32362.
PT   New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT   - useful for diagnosis of M. tuberculosis infection
PS   Claim 11; Page 59-60; 190pp; English.
CC   A new immunogenic polypeptide has been developed comprising an
CC   immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC   its variant differing only in conservative substitutions and/or
CC   modifications). The present sequence encodes a specifically claimed
CC   M.tuberculosis antigen, TbRa28. The immunogenic polypeptide can be
CC   used to diagnose M.tuberculosis infection by forming complexes with
CC   specific antibodies in the sample. Fragments of DNA encoding the
CC   immunogenic polypeptide can be used as diagnostic primers or probes
CC   and agents that bind to the antigen, especially monoclonal antibodies
CC   or equivalent polyclonal antibodies, are also used for diagnosis.
SQ   Sequence   1155 BP;    203 A;    377 C;    394 G;    181 T;


  Query Match             88.9%;  Score 16;  DB 1;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATC 16
          ||||||||||||||||
Db    806 CGCGGCCGACATCATC 821


RESULT   8
V44350
ID   V44350 standard; DNA; 1155 BP.
AC   V44350;
DT   09-NOV-1998  (first entry)
DE   Mycobacterium tuberculosis antigen TbRa28 DNA.
KW   Tuberculosis; infection; diagnosis; antigen; TbRa28; ss.
OS   Mycobacterium tuberculosis strain H37Ra.
FH   Key             Location/Qualifiers
FT   CDS             2. .928
FT                   /*tag= a
PN   WO9816645-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18214.
PR   13-MAR-1997; US-818111.
PR   11-OCT-1996; US-729622.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-251292/22.
DR   P-PSDB; W64302.
PT   New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
```

```
PT   and diagnosis of tuberculosis
PS   Claim 11a; Page 67-68; 250pp; English.
CC   This DNA sequence codes for Mycobacterium tuberculosis soluble
CC   antigen TbRa28 (see W64302). It was isolated from a M. tuberculosis
CC   strain H37Ra expression library with rabbit anti-sera raised
CC   against M. tuberculosis supernatant. TbRa28 has previously been
CC   identified in M. tuberculosis. The invention relates to
CC   compositions and methods for diagnosing tuberculosis. It provides
CC   polypeptides (see W64291-W64379) comprising an antigenic portion of
CC   a soluble M. tuberculosis antigen, or an immunogenic portion of an
CC   M. tuberculosis antigen, as well as DNA sequences encoding such
CC   polypeptides, recombinant expression vectors and transformed or
CC   transfected host cells. Also claimed are methods and diagnostic
CC   kits for detecting M. tuberculosis infection in a patient using
CC   these polypeptides, antibodies or oligonucleotide probes and
CC   primers, for the diagnosis of tuberculosis.
SQ   Sequence   1155 BP;    203 A;    377 C;    394 G;    181 T;


  Query Match            88.9%;  Score 16;  DB 1;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATC 16
          ||||||||||||||||
Db    806 CGCGGCCGACATCATC 821


RESULT   9
V64458
ID   V64458 standard; DNA; 1155 BP.
AC   V64458;
DT   27-JAN-1999  (first entry)
DE   M. tuberculosis immunogenic polypeptide TbRa28 DNA.
KW   Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW   vaccine; pharmaceutical; infection; diagnosis; ss.
OS   Mycobacterium tuberculosis.
PN   WO9816646-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18293.
PR   13-MAR-1997; US-818112.
PR   11-OCT-1996; US-730510.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-261042/23.
DR   P-PSDB; W81665.
PT   Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and for diagnosis, treatment and prevention of tuberculosis
PS   Claim 11; Page 68-69; 230pp; English.
CC   This sequence encodes an immunogenic portion of a soluble Mycobacterium
CC   tuberculosis (MT) antigen which can be used in a method for inducing
CC   protective immunity against tuberculosis (TB). This sequence can be
CC   formulated into vaccines and/or pharmaceutical compositions for
CC   immunising against M. tuberculosis infection or may be used for the
CC   diagnosis of tuberculosis.
SQ   Sequence   1155 BP;    203 A;    377 C;    394 G;    181 T;


  Query Match            88.9%;  Score 16;  DB 1;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCGGCCGACATCATC 16
          ||||||||||||||||
Db    806 CGCGGCCGACATCATC 821


RESULT  10
X18867
ID   X18867 standard; DNA; 12019 BP.
AC   X18867;
DT   10-MAY-1999  (first entry)
DE   Alcaligenes sp. Pox proteins R, A, B, C, D, E, F, G, H and I coding DNA.
KW   Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;
KW   PoxI; aromatic; oxidative; petroleum purification; chemical industry;
KW   drug; ds.
OS   Alcaligenes sp.
PN   J11042088-A.
PD   16-FEB-1999.
PF   25-JUL-1997; 200625.
PR   25-JUL-1997; JP-200625.
PA   (TOFU ) TONEN CORP.
DR   WPI; 99-197820/17.
DR   P-PSDB; W98969, W98970, W98971, W98972, W98973, W98974, W98975,
DR          W98976, W98977, W98978.
PT   New encoding an aromatic cpd. oxidative decompsn. enzyme - useful in
PT   the fields of petroleum purification, chemical and drug industries
PS   Claim 1; Page 7-19; 35pp; Japanese.
CC   The present sequence encodes Alcaligenes sp. proteins PoxR, PoxA, PoxB,
CC   PoxC, PoxD, PoxE, PoxF, PoxG, PoxH and PoxI. The proteins are useful in
CC   the fields of petroleum purification, chemical industry and drug
CC   industry related to the synthesis, conversion and decomposition of
CC   aromatic compounds.
SQ   Sequence   12019 BP;    2015 A;    4104 C;    4151 G;    1747 T;


  Query Match            85.6%;  Score 15.4;  DB 1;  Length 12019;
  Best Local Similarity  94.1%;  Pred. No. 42;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 GCGGCCGACATCATCGC 18
          |||||||||||||||| |
Db  10290 GCGGCCGACATCATCTC 10306


RESULT  11
Q13288
ID   Q13288 standard; DNA; 13144 BP.
AC   Q13288;
DT   25-OCT-1991  (first entry)
DE   P.denitrificans genes cob q, p, w, n and o.
KW   cob gene; corrinoid; descobaltocorrinoid; cor gene; ds.
OS   Pseudomonas denitrificans.
FH   Key             Location/Qualifiers
FT   cds             429. .1886
FT                   /*tag= a
FT                   /product= COBQ
FT                   /note= "has cobyric acid synthase activity"
FT   cds             3364. .3888
FT                   /*tag= b
FT                   /product= COBP
FT                   /note= "involved in conversion of cobinamide to
FT                   GDP-cobinamide; has cobinamide kinase
FT                   and cobinamide phosphate guanyltransferase
FT                   activity"
FT   cds             3892. .4956
FT                   /*tag= c
FT                   /product= COBW
FT   cds             5060. .8887
FT                   /*tag= d
FT                   /product= COBN
FT   cds             9034. .9678
FT                   /*tag= e
FT                   /product= COBO
FT                   /note= "has cob(I)alamin adenosyltransferase
FT                   activity"
PN   WO9111518-A.
PD   08-AUG-1991.
PF   30-JAN-1991; F00054.
PR   31-JAN-1990; FR-001137.
PA   (RHON ) RHONE-POULENC BIOCH.
PI   Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;
PI   Thibaut D;
DR   WPI; 91-252650/34.
```

```
DR   P-PSDB; R13513-R13517.
PT   New polypeptide(s) involved in cobalamin and cobamide
PT   biosynthesis - and DNA encoding them, for amplification of
PT   cobalamin, esp. coenzyme B12 prodn.
PS   Claim 1; Fig 33; 299pp; French.
CC   This is a P.denitrificans genomic fragment isolated from a DNA bank
CC   constructed in vector pXL59. Plasmids able to complement P.putida and
CC   Agrobacterium tumefaciens cob mutants were identified and their
CC   inserts were sequenced. This insert was found to contain 5 putative
CC   open reading frames. The presence of cob genes n, o, p, q and w
CC   was subsequently verified by genetic analysis.
CC   See also Q13284-Q13287.
SQ   Sequence   13144 BP;    2489 A;    4078 C;    4291 G;    2286 T;


  Query Match            85.6%;  Score 15.4;  DB 1;  Length 13144;
  Best Local Similarity  94.1%;  Pred. No. 42;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 GCGGCCGACATCATCGC 18
          |||||||||||||  |||
Db  12893 GCGGCCGACATCAGCGC 12909


RESULT  12
V49523
ID   V49523 standard; DNA; 20 BP.
AC   V49523;
DT   20-OCT-1998  (first entry)
DE   Mycobacterium sp. AlaDH oligonucleotide AlaDH-R5.
KW   Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW   swimmers disease; vaccine; epidemic; infection; identification; ss.
OS   Synthetic.
OS   Mycobacterium sp.
PN   WO9832862-A2.
PD   30-JUL-1998.
PF   29-JAN-1998; E00484.
PR   29-JAN-1997; EP-101339.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-427958/36.
PT   Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT   - used for diagnosis of tuberculosis and other mycobacterial
PT   diseases, also for treatment and prevention, for drug screening and
PT   for bio-transformation
PS   Disclosure; Page 10; 57pp; German.
CC   V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC   dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC   to diagnose tuberculosis and other mycobacterial infections (including
CC   'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC   animals. The protein can also be used for control of epidemics and for
CC   vaccination, to screen for agents with anti-mycobacterial activity, and
CC   in bio-transformations that are specific for L-alanine. Also mycobacteria
CC   can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC   that is secreted early during infection.
SQ   Sequence   20 BP;    3 A;    9 C;    4 G;    4 T;


  Query Match            83.3%;  Score 15;  DB 1;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 67;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GGCCGACATCATCGC 18
          |||||||||||||||
Db      1 GGCCGACATCATCGC 15


RESULT  13
V49621
ID   V49621 standard; DNA; 20 BP.
AC   V49621;
DT   20-NOV-1998  (first entry)
DE   AlaDH derived oligonucleotide AlaDH-R5.
KW   ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS   Mycobacterium sp.
PN   WO9836089-A2.
PD   20-AUG-1998.
PF   29-JAN-1998; E00483.
PR   29-JAN-1997; EP-101338.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-457123/39.
PT   Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT   - useful for, e.g. for diagnosis, differentiation of strains,
PT   monitoring vaccination and identification of mycobacterial
PT   inhibitors
PS   Claim 9; Fig 2.5; 55pp; German.
CC   THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC   are used for diagnosing tuberculosis (TB) and other mycobacterial
CC   infections in humans or animals.  Kits are used for direct diagnosis of
CC   TB on clinical samples (e.g. body fluids) and can differentiate between
CC   pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC   the M.tuberculosis (M. t.) complex.  The kit may also be used to
CC   identify substances that inhibit mycobacteria, for combatting epidemics
CC   and for vaccination follow-up.  The oligonucleotides are used similarly
CC   in diagnostic hybridisation tests, also for culture confirmation of
CC   isolated strains and for chromosome fingerprinting to detect/
CC   differentiate between mycobacteria, and for L-alanine-specific
CC   biotransformation reactions.  AlaDH is an early antigen, present
CC   extracellularly after only a few days of growth, making it an ideal
CC   drug target.
SQ   Sequence   20 BP;    3 A;    9 C;    4 G;    4 T;


  Query Match            83.3%;  Score 15;  DB 1;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 67;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GGCCGACATCATCGC 18
          |||||||||||||||
Db      1 GGCCGACATCATCGC 15


RESULT  14
V59465
ID   V59465 standard; DNA; 2664 BP.
AC   V59465;
DT   24-DEC-1998  (first entry)
DE   Nucleotide sequence encoding Mycobacterium Kansasii polymerase.
KW   ss; polymerase; Mycobacterium; diagnosis; treatment.
OS   Mycobacterium Kansasii.
FH   Key             Location/Qualifiers
FT   CDS             1. .2664
FT                   /*tag= a
FT                   /product= "Mycobacterium kansasii polymerase"
PN   J10248583-A.
PD   22-SEP-1998.
PF   17-MAR-1997; 083242.
PR   17-MAR-1997; JP-083242.
PA   (BMLB-) BML KK.
DR   WPI; 98-560735/48.
DR   P-PSDB; W77409.
PT   DNA polymerase and gene encoding it - useful in diagnosis and.
PT   treatment
PS   Claim 6; Fig 3; 16pp; Japanese.
CC   The DNA polymerase and the DNA polymerase gene derived from a
CC   microbe of the Mycobacterium genus are useful in diagnosis and
CC   treatments.
SQ   Sequence   2664 BP;    466 A;    914 C;    890 G;    394 T;


  Query Match            83.3%;  Score 15;  DB 1;  Length 2664;
  Best Local Similarity  100.0%;  Pred. No. 65;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GCGGCCGACATCATC 16
```

```
          ||||||||||||||||
Db    2428 GCGGCCGACATCATC 2442


RESULT  15
Q55756/c
ID   Q55756 standard; DNA; 6317 BP.
AC   Q55756;
DT   26-SEP-1995  (first entry)
DE   Enterobacter cloacae genomic probe ET-12.
KW   Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW   K.pneumoniae; E.cloacae; clinical sample; ds.
OS   Enterobacter cloacae.
PN   WO9401583-A.
PD   20-JAN-1994.
PF   07-JUL-1993; J00936.
PR   07-JUL-1992; JP-179719.
PA   (FUSO ) FUSO PHARM IND LTD.
PA   (OHNO/) OHNO T.
PI   Eda S,  Matsuhisa A,  Ohno T,  Uehara H;
DR   WPI; 94-035086/04.
PT   Probe for identifying bacteria causing infectious disease -
PT   consists of a DNA fragment obtained by HindIII cleavage of the
PT   pathogenic bacterial genomic DNA
PS   Claim 7; Page 74-77; 100pp; Japanese.
CC   The nucleotide sequence of a 6317 bp probe obtained by digestion of
CC   Enterobacter cloacae genomic DNA with the restriction enzyme HindIII.
CC   The probes (Q55133-48 and Q55752-58) represent other probes derived,
CC   by HindIII digestion of the genomes of Staphylococcus aureus,
CC   S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC   Klebsiella pneumoniae or Enterobacter cloacae.  The probes can be used
CC   to detect their respective microorganisms in clinical samples.
SQ   Sequence   6317 BP;    1420 A;    1758 C;    1810 G;    1329 T;


  Query Match             83.3%;  Score 15;  DB 1;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATC 16
          ||||||||||||||||
Db    6178 GCGGCCGACATCATC 6164


Search completed: June 22, 2000, 18:01:44
Job time: 10116 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:22:13 ; Search time 123.79 Seconds
(without alignments)
18.901 Million cell updates/sec

Title: US-09-362-485-21
Perfect score: 18
Sequence: 1 CGCGGCCGACATCATCGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	15	83.3	6317	2	US-08-920-812-21	Sequence 21, Appl
c	2	15	83.3	6317	2	US-08-920-827-21	Sequence 21, Appl
c	3	15	83.3	6317	2	US-08-921-177-21	Sequence 21, Appl
c	4	15	83.3	6317	2	US-08-362-577C-21	Sequence 21, Appl
c	5	15	83.3	6317	3	US-08-920-828-21	Sequence 21, Appl
c	6	14.8	82.2	1185	4	US-08-628-039-9	Sequence 9, Appli
	7	14.8	82.2	3252	4	US-08-809-740A-1	Sequence 1, Appli
	8	14.8	82.2	3252	4	US-08-809-740A-4	Sequence 4, Appli
c	9	14.4	80.0	11219	2	US-07-642-734C-1	Sequence 1, Appli
	10	14	77.8	4131	1	US-08-485-588-4	Sequence 4, Appli
	11	14	77.8	4131	2	US-08-484-565-4	Sequence 4, Appli
	12	14	77.8	4131	3	US-08-480-751-4	Sequence 4, Appli
	13	14	77.8	4131	4	US-08-943-986-4	Sequence 4, Appli
	14	14	77.8	4131	5	US-08-353-784-4	Sequence 4, Appli
	15	13.8	76.7	84	4	US-08-426-599B-10	Sequence 10, Appl
	16	13.8	76.7	624	4	US-08-426-599B-3	Sequence 3, Appli
	17	13.8	76.7	759	4	US-08-853-659A-18	Sequence 18, Appl
c	18	13.8	76.7	1621	4	US-08-933-750C-96	Sequence 96, Appl
	19	13.8	76.7	1879	3	US-08-403-852D-5	Sequence 5, Appli
c	20	13.8	76.7	2214	1	US-07-985-458-1	Sequence 1, Appli
	21	13.8	76.7	2574	3	US-08-677-734A-8	Sequence 8, Appli
	22	13.8	76.7	2744	5	US-09-071-101-1	Sequence 1, Appli
	23	13.8	76.7	2800	5	US-08-448-194-3	Sequence 3, Appli
	24	13.8	76.7	3237	5	US-08-451-946B-5	Sequence 5, Appli
	25	13.8	76.7	3237	5	US-08-446-938B-5	Sequence 5, Appli
	26	13.8	76.7	3237	5	US-08-311-703A-5	Sequence 5, Appli
	27	13.8	76.7	3237	5	US-08-446-939B-5	Sequence 5, Appli
	28	13.8	76.7	3237	5	PCT-US92-09326-1	Sequence 1, Appli
	29	13.8	76.7	3537	4	US-08-363-124A-3	Sequence 3, Appli
	30	13.8	76.7	6122	1	US-08-403-545-1	Sequence 1, Appli
	31	13.8	76.7	15512	4	US-08-853-659A-5	Sequence 5, Appli
c	32	13.8	76.7	15512	4	US-08-853-659A-8	Sequence 8, Appli
	33	13.8	76.7	15512	4	US-08-853-659A-63	Sequence 63, Appl
c	34	13.8	76.7	15512	4	US-08-853-659A-66	Sequence 66, Appl
	35	13.8	76.7	24701	4	US-08-853-659A-2	Sequence 2, Appli
c	36	13.8	76.7	24701	4	US-08-853-659A-3	Sequence 3, Appli
	37	13.8	76.7	24701	4	US-08-853-659A-60	Sequence 60, Appl
c	38	13.8	76.7	24701	4	US-08-853-659A-61	Sequence 61, Appl
	39	13.8	76.7	30001	1	US-08-125-468-1	Sequence 1, Appli
	40	13.8	76.7	30001	3	US-08-474-933-1	Sequence 1, Appli
c	41	13.8	76.7	49272	2	US-08-614-770A-1	Sequence 1, Appli
	42	13.4	74.4	890	5	US-09-057-969-12	Sequence 12, Appl
	43	13.4	74.4	1142		US-09-057-969-11	Sequence 11, Appl
	44	13.4	74.4	1197	5	PCT-US94-14277-7	Sequence 7, Appli
	45	13.4	74.4	1379	5	US-09-057-969-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-08-920-812-21/c
; Sequence 21, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter cloacae
;      STRAIN:  Clinical Isolate ET-12
US-08-920-812-21
```

```
  Query Match             83.3%;  Score 15;  DB 2;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATC 16
           ||||||||||||||||
Db    6178 GCGGCCGACATCATC 6164


RESULT   2
US-08-920-827-21/c
; Sequence 21, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter cloacae
;      STRAIN:  Clinical Isolate ET-12
US-08-920-827-21


  Query Match             83.3%;  Score 15;  DB 2;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATC 16
           ||||||||||||||||
Db    6178 GCGGCCGACATCATC 6164


RESULT   3
US-08-921-177-21/c
; Sequence 21, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter cloacae
;      STRAIN:  Clinical Isolate ET-12
US-08-921-177-21


  Query Match             83.3%;  Score 15;  DB 2;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 GCGGCCGACATCATC 16
           ||||||||||||||||
Db    6178 GCGGCCGACATCATC 6164


RESULT   4
US-08-362-577C-21/c
; Sequence 21, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
```

```
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/362,577C
;       FILING DATE:  27-MAR-1995
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Rin-Laures, Li-Hsien
;       REGISTRATION NUMBER:  33,547
;       REFERENCE/DOCKET NUMBER:  19036/32420
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6300
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6317 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Enterobacter cloacae
;       STRAIN:  Clinical Isolate ET-12
US-08-362-577C-21


  Query Match             83.3%;  Score 15;  DB 2;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GCGGCCGACATCATC 16
          |||||||||||||||
Db   6178 GCGGCCGACATCATC 6164


RESULT   5
US-08-920-828-21/c
; Sequence 21, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter cloacae
;      STRAIN:  Clinical Isolate ET-12
US-08-920-828-21


  Query Match             83.3%;  Score 15;  DB 3;  Length 6317;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GCGGCCGACATCATC 16
          |||||||||||||||
Db   6178 GCGGCCGACATCATC 6164


RESULT   6
US-08-628-039-9/c
; Sequence 9, Application US/08628039
; Patent No. 5942660
;  GENERAL INFORMATION:
;    APPLICANT:  Guys, Kenneth J.
;    APPLICANT:  Mitsky, Timothy A.
;    APPLICANT:  Kishore, Ganesh M.
;    APPLICANT:  Slater, Steven C.
;    APPLICANT:  Padgette, Stephen R.
;    APPLICANT:  Stark, David M.
;    APPLICANT:  Hinchee, Maud A. W.
;    APPLICANT:  Clemente, Thomas E.
;    APPLICANT:  Connor-Ward, Dannette V.
;    APPLICANT:  Fedele, Mary J.
;    APPLICANT:  Fry, Joyce E.
;    APPLICANT:  Howe, Arlene R.
;    APPLICANT:  Rozman, Renee J.
;    TITLE OF INVENTION:  Methods of Optimizing Substrate Pools and
;    TITLE OF INVENTION:  Biosynthesis of Poly-B-hydroxybutyrate-co-poly-B-hydroxyval
;    TITLE OF INVENTION:  in Bacteria and Plants
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gary M. Bond, Monsanto Company, A3SB
;      STREET:  800 No. 5942660th Lindbergh Boulevard
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  USA
;      ZIP:  63167
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,039
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:  13-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bond, Gary
;      REGISTRATION NUMBER:  29,283
;      REFERENCE/DOCKET NUMBER:  38-21(13585)A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)694-3412
;      TELEFAX:  (314)695-5435
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1185 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-628-039-9


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 1185;
  Best Local Similarity   88.9%;  Pred. No. 60;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           || |||||||||| ||||
Db     154 CGTGGCCGACATCGTCGC 137


RESULT   7
US-08-809-740A-1
; Sequence 1, Application US/08809740A
; Patent No. 5986077
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Process for producing anthracyclines
;    TITLE OF INVENTION:  and intermediates thereof
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Evenson, McKeown, Edwards and Lenahan
;      STREET:  1200 G Street, Suite 700
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/809,740A
;      FILING DATE:  27-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI95/00537
;      FILING DATE:  30-SEP-1995
;      APPLICATION NUMBER:  FI 944556
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  H. Thomas Anderton, Jr.
;      REGISTRATION NUMBER:  40,895
;      REFERENCE/DOCKET NUMBER:  1574/43419
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      STRAIN:  Streptomyces nogalater ATCC 27451
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  359..1651
;      OTHER INFORMATION:  /note= "ORF1"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2937..3197
;      OTHER INFORMATION:  /note= "ORF3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1648..1651
;      OTHER INFORMATION:  /note= "overlapping sequence in
;      OTHER INFORMATION:  ORF1 and ORF2"
US-08-809-740A-1


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 3252;
  Best Local Similarity   88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCGGCCGACATCATCGC 18
           |||||||||| | |||||
Db     907 CGCGGCCGACCTGATCGC 924


RESULT   8
US-08-809-740A-4
; Sequence 4, Application US/08809740A
; Patent No. 5986077
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Process for producing anthracyclines
;    TITLE OF INVENTION:  and intermediates thereof
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Evenson, McKeown, Edwards and Lenahan
;      STREET:  1200 G Street, Suite 700
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/809,740A
;      FILING DATE:  27-MAR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI95/00537
;      FILING DATE:  30-SEP-1995
;      APPLICATION NUMBER:  FI 944556
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  H. Thomas Anderton, Jr.
;      REGISTRATION NUMBER:  40,895
;      REFERENCE/DOCKET NUMBER:  1574/43419
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      STRAIN:  Streptomyces nogalater ATCC 27451
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1648..2877
;      OTHER INFORMATION:  /note= "ORF2"
US-08-809-740A-4


  Query Match             82.2%;  Score 14.8;  DB 4;  Length 3252;
  Best Local Similarity   88.9%;  Pred. No. 56;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 CGCGGCCGACATCATCGC 18
           |||||||||| | |||||
Db     907 CGCGGCCGACCTGATCGC 924


RESULT   9
US-07-642-734C-1/c
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9907..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1


  Query Match             80.0%;  Score 14.4;  DB 2;  Length 11219;
  Best Local Similarity   93.8%;  Pred. No. 79;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CGGCCGACATCATCGC 18
           ||||||||| ||||||
Db    5271 CGGCCGACACCATCGC 5256


RESULT  10
US-08-485-588-4
; Sequence 4, Application US/08485588
; Patent No. 5688938
;  GENERAL INFORMATION:
;    APPLICANT:  Edward M. Brown
;    APPLICANT:  Steven C. Hebert
;    APPLICANT:  Forrest H. Fuller
;    APPLICANT:  James E. Garrett, Jr.
;    TITLE OF INVENTION:  CALCIUM RECEPTOR-ACTIVE
;    TITLE OF INVENTION:  MOLECULES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  First Interstate World Center
```

```
;       STREET:  Suite 4700
;       STREET:  633 West Fifth Street
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  FASTSEQ
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/485,588
;       FILING DATE:  7 June, 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:  9
;       APPLICATION NUMBER:  08/353,784
;       FILING DATE:  9 December, 1994
;       APPLICATION NUMBER:  PCT/US/94/12117
;       FILING DATE:  21 October, 1994
;       APPLICATION NUMBER:  U.S. 08/292,827
;       FILING DATE:  23 August, 1994
;       APPLICATION NUMBER:  U.S. 08/141,248
;       FILING DATE:  22 October, 1993
;       APPLICATION NUMBER:  U.S. 08/009,389
;       FILING DATE:  23 February, 1993
;       APPLICATION NUMBER:  U.S. 08/017,127
;       FILING DATE:  12 February, 1993
;       APPLICATION NUMBER:  U.S. 07/934,161
;       FILING DATE:  21 August, 1992
;       APPLICATION NUMBER:  U.S. 07/834,044
;       FILING DATE:  11 February, 1992
;       APPLICATION NUMBER:  U.S. 07/749,451
;       FILING DATE:  23 August, 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Heber, Sheldon O.
;       REGISTRATION NUMBER:  38,179
;       REFERENCE/DOCKET NUMBER:  213/005
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4131 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  574..3810
;       OTHER INFORMATION:
US-08-485-588-4


  Query Match           77.8%;  Score 14;  DB 1;  Length 4131;
  Best Local Similarity  100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GGCCGACATCATCG 17
           ||||||||||||||
Db    1164 GGCCGACATCATCG 1177


RESULT  11
US-08-484-565-4
; Sequence 4, Application US/08484565
; Patent No. 5763569
;  GENERAL INFORMATION:
;    APPLICANT:  Edward M. Brown
;    APPLICANT:  Steven C. Hebert
;    APPLICANT:  James E. Garrett, Jr.
;    TITLE OF INVENTION:  CALCIUM RECEPTOR-ACTIVE
;    TITLE OF INVENTION:  MOLECULES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  First Interstate World Center
;      STREET:  Suite 4700
;      STREET:  633 West Fifth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FASTSEQ
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,565
;      FILING DATE:  7 June, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:  9
;      APPLICATION NUMBER:  08/353,784
;      FILING DATE:  9 December, 1994
;      APPLICATION NUMBER:  PCT/US/94/12117
;      FILING DATE:  21 October, 1994
;      APPLICATION NUMBER:  U.S. 08/292,827
;      FILING DATE:  23 August, 1994
;      APPLICATION NUMBER:  U.S. 08/141,248
;      FILING DATE:  22 October, 1993
;      APPLICATION NUMBER:  U.S. 08/009,389
;      FILING DATE:  23 February, 1993
;      APPLICATION NUMBER:  U.S. 08/017,127
;      FILING DATE:  12 February, 1993
;      APPLICATION NUMBER:  U.S. 07/934,161
;      FILING DATE:  21 August, 1992
;      APPLICATION NUMBER:  U.S. 07/834,044
;      FILING DATE:  11 February, 1992
;      APPLICATION NUMBER:  U.S. 07/749,451
;      FILING DATE:  23 August, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heber, Sheldon O.
;      REGISTRATION NUMBER:  38,179
;      REFERENCE/DOCKET NUMBER:  213/006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4131 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  574..3810
;      OTHER INFORMATION:
US-08-484-565-4


  Query Match           77.8%;  Score 14;  DB 2;  Length 4131;
  Best Local Similarity  100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GGCCGACATCATCG 17
           ||||||||||||||
Db    1164 GGCCGACATCATCG 1177
```

```
RESULT  12
US-08-480-751-4
; Sequence 4, Application US/08480751
; Patent No. 5858684
;  GENERAL INFORMATION:
;    APPLICANT:  Edward F. Nemeth
;    APPLICANT:  Edward M. Brown
;    APPLICANT:  Steven C. Hebert
;    APPLICANT:  Forrest H. Fuller
;    APPLICANT:  James E. Garrett, Jr.
;    TITLE OF INVENTION:  CALCIUM RECEPTOR-ACTIVE
;    TITLE OF INVENTION:  MOLECULES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  First Interstate World Center
;      STREET:  Suite 4700
;      STREET:  633 West Fifth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FASTSEQ
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,751
;      FILING DATE:  7 June, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:  9
;      APPLICATION NUMBER:  08/353,784
;      FILING DATE:  9 December, 1994
;      APPLICATION NUMBER:  PCT/US/94/12117
;      FILING DATE:  21 October, 1994
;      APPLICATION NUMBER:  U.S. 08/292,827
;      FILING DATE:  23 August, 1994
;      APPLICATION NUMBER:  U.S. 08/141,248
;      FILING DATE:  22 October, 1993
;      APPLICATION NUMBER:  U.S. 08/009,389
;      FILING DATE:  23 February, 1993
;      APPLICATION NUMBER:  U.S. 08/017,127
;      FILING DATE:  12 February, 1993
;      APPLICATION NUMBER:  U.S. 07/934,161
;      FILING DATE:  21 August, 1992
;      APPLICATION NUMBER:  U.S. 07/834,044
;      FILING DATE:  11 February, 1992
;      APPLICATION NUMBER:  U.S. 07/749,451
;      FILING DATE:  23 August, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heber, Sheldon O.
;      REGISTRATION NUMBER:  38,179
;      REFERENCE/DOCKET NUMBER:  213/004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4131 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  574..3810
;      OTHER INFORMATION:
```

```
US-08-480-751-4


  Query Match             77.8%;  Score 14;  DB 3;  Length 4131;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GGCCGACATCATCG 17
           ||||||||||||||
Db    1164 GGCCGACATCATCG 1177


RESULT  13
US-08-943-986-4
; Sequence 4, Application US/08943986
; Patent No. 5962314
;  GENERAL INFORMATION:
;    APPLICANT:  Edward M. Brown
;    APPLICANT:  Steven C. Hebert
;    APPLICANT:  James E. Garrett, Jr.
;    TITLE OF INVENTION:  CALCIUM RECEPTOR-ACTIVE
;    TITLE OF INVENTION:  MOLECULES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  First Interstate World Center
;      STREET:  Suite 4700
;      STREET:  633 West Fifth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FASTSEQ
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/943,986
;      FILING DATE:  03-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/484,565
;      FILING DATE:  7-June-1995
;      APPLICATION NUMBER:  08/353,784
;      FILING DATE:  9 December, 1994
;      APPLICATION NUMBER:  PCT/US/94/12117
;      FILING DATE:  21 October, 1994
;      APPLICATION NUMBER:  U.S. 08/292,827
;      FILING DATE:  23 August, 1994
;      APPLICATION NUMBER:  U.S. 08/141,248
;      FILING DATE:  22 October, 1993
;      APPLICATION NUMBER:  U.S. 08/009,389
;      FILING DATE:  23 February, 1993
;      APPLICATION NUMBER:  U.S. 08/017,127
;      FILING DATE:  12 February, 1993
;      APPLICATION NUMBER:  U.S. 07/934,161
;      FILING DATE:  21 August, 1992
;      APPLICATION NUMBER:  U.S. 07/834,044
;      FILING DATE:  11 February, 1992
;      APPLICATION NUMBER:  U.S. 07/749,451
;      FILING DATE:  23 August, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heber, Sheldon O.
;      REGISTRATION NUMBER:  38,179
;      REFERENCE/DOCKET NUMBER:  213/006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4131 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  574..3810
;      OTHER INFORMATION:
US-08-943-986-4


  Query Match             77.8%;  Score 14;  DB 4;  Length 4131;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GGCCGACATCATCG 17
           ||||||||||||||
Db    1164 GGCCGACATCATCG 1177


RESULT  14
US-08-353-784-4
; Sequence 4, Application US/08353784
; Patent No. 6011068
;  GENERAL INFORMATION:
;    APPLICANT:  Edward F. Nemeth, Edward M.
;    APPLICANT:  Brown, Steven C. Hebert,
;    APPLICANT:  Bradford C. Van Wagenen, Manuel
;    APPLICANT:  F. Balandrin, Forrest H. Fuller,
;    APPLICANT:  Eric G. DelMar, and Scott T. Moe
;    TITLE OF INVENTION:  CALCIUM RECEPTOR-ACTIVE
;    TITLE OF INVENTION:  MOLECULES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  First Interstate World Center
;      STREET:  Suite 4700
;      STREET:  633 West Fifth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FASTSEQ
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,784
;      FILING DATE:  9 December, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:  8
;      APPLICATION NUMBER:  PCT/US/94/12117
;      FILING DATE:  21 October, 1994
;      APPLICATION NUMBER:  U.S. 08/292,827
;      FILING DATE:  23 August, 1994
;      APPLICATION NUMBER:  U.S. 08/141,248
;      FILING DATE:  22 October, 1993
;      APPLICATION NUMBER:  U.S. 08/009,389
;      FILING DATE:  23 February, 1993
;      APPLICATION NUMBER:  U.S. 08/017,127
;      FILING DATE:  12 February, 1993
;      APPLICATION NUMBER:  U.S. 07/934,161
;      FILING DATE:  21 August, 1992
;      APPLICATION NUMBER:  U.S. 07/834,044
;      FILING DATE:  11 February, 1992
;      APPLICATION NUMBER:  U.S. 07/749,451
;      FILING DATE:  23 August, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heber, Sheldon O.
;      REGISTRATION NUMBER:  38,179
;      REFERENCE/DOCKET NUMBER:  209/069
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4131 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  574..3810
;      OTHER INFORMATION:
US-08-353-784-4


  Query Match             77.8%;  Score 14;  DB 5;  Length 4131;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GGCCGACATCATCG 17
           ||||||||||||||
Db    1164 GGCCGACATCATCG 1177


RESULT  15
US-08-426-599B-10
; Sequence 10, Application US/08426599B
; Patent No. 5932438
;  GENERAL INFORMATION:
;    APPLICANT:  Uriach-Marsal, Juan
;    APPLICANT:  Rubio-Susan, Victor
;    APPLICANT:  Patino-Martin, Cristina
;    APPLICANT:  Kalo-Koenova, Eliza I.
;    APPLICANT:  del Moral-Juarez, Catalina
;    APPLICANT:  Faus-Santasusana, Ignacio
;    APPLICANT:  del Rio-Pericacho, Jose-Luis
;    APPLICANT:  Blade-Pique, Joan
;    TITLE OF INVENTION:  Preparation of Thaumatin Sweeteners
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rothwell, Figg, Ernst & Kurz, P.C.
;      STREET:  555 Thirteenth Street, N.W., Suite 701 East
;      STREET:  Tower
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,599B
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saxe, Stephen A.
;      REGISTRATION NUMBER:  38,609
;      REFERENCE/DOCKET NUMBER:  1604-123A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-624-1589
;      TELEFAX:  202-783-6031
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  84 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

; MOLECULE TYPE: other nucleic acid
US-08-426-599B-10

Query Match 76.7%; Score 13.8; DB 4; Length 84;
Best Local Similarity 88.2%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 CGCGGCCGACATCATCG 17
           ||| ||||||||| |||
Db      54 CGCCGCCGACATCGTCG 70
```

Search completed: June 22, 2000, 18:03:58
Job time: 9705 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:26 ; Search time 5541.94 Seconds
(without alignments)
12.433 Million cell updates/sec

Title: US-09-362-485-16
Perfect score: 17
Sequence: 1 GGGGCCGTCCTGGTGCC 17

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	17	100.0	575	63	AI951640	AI951640 wv37f09.x
	2	16	94.1	275	69	AW124365	AW124365 UI-M-BH2.
c	3	15.4	90.6	246	40	AA911158	AA911158 oe69e07.s
	4	15.4	90.6	293	61	AI836163	AI836163 UI-M-AQ0-
c	5	15.4	90.6	333	102	AQ163728	AQ163728 HS_2247_B
	6	15.4	90.6	388	45	AI377512	AI377512 tc37d11.x
c	7	15.4	90.6	430	102	AQ143042	AQ143042 HS_3064_A
c	8	15.4	90.6	476	79	AW260226	AW260226 um75a09.x
	9	15.4	90.6	630	35	AA544009	AA544009 vj83h11.r
	10	15	88.2	295	34	AA518462	AA518462 vi06h11.r
c	11	15	88.2	401	92	AQ911288	AQ911288 GSSTc0332
c	12	15	88.2	483	41	AI004179	AI004179 ou56c12.x
	13	15	88.2	536	74	AW328031	AW328031 dr03g02.x
	14	15	88.2	636	61	AL041617	AL041617 DKFZp434K
	15	14.4	84.7	106	63	AI950970	AI950970 wx61f11.x
	16	14.4	84.7	174	42	AI081643	AI081643 ou63a07.s
c	17	14.4	84.7	180	35	AA591969	AA591969 vi50c07.r
c	18	14.4	84.7	190	28	AA090821	AA090821 yy0193.se
	19	14.4	84.7	205	32	AA339541	AA339541 EST44999
	20	14.4	84.7	226	23	R89660	R89660 ym97e04.r1
	21	14.4	84.7	231	62	AI892075	AI892075 mg99g05.y
	22	14.4	84.7	232	80	AW404312	AW404312 UI-HF-BL0
	23	14.4	84.7	236	35	AA583313	AA583313 nn37g10.s
	24	14.4	84.7	248	20	Z39051	Z39051 HSC0YC012 n
	25	14.4	84.7	250	49	AI655096	AI655096 wb67a12.x
	26	14.4	84.7	263	48	AI571084	AI571084 to26f11.x
	27	14.4	84.7	280	21	F08461	F08461 HSCZUH031 n
	28	14.4	84.7	283	74	AW194360	AW194360 xm08f01.x
	29	14.4	84.7	284	48	AI560534	AI560534 tq47g12.x
c	30	14.4	84.7	288	81	AW428401	AW428401 66347 MAR
	31	14.4	84.7	289	23	H30880	H30880 yo79g12.r1
c	32	14.4	84.7	295	31	AA327118	AA327118 EST30431
	33	14.4	84.7	295	79	AW294467	AW294467 UI-H-BI2-
	34	14.4	84.7	302	32	AA355927	AA355927 EST64400
	35	14.4	84.7	313	20	T34152	T34152 EST63414 Hu
	36	14.4	84.7	313	31	AA316749	AA316749 EST18842
	37	14.4	84.7	313	80	AW406984	AW406984 UI-HF-BL0
	38	14.4	84.7	319	21	R07175	R07175 yf14c04.s1
	39	14.4	84.7	327	50	AI680300	AI680300 tw62b06.x
	40	14.4	84.7	337	79	AW290972	AW290972 UI-H-BI2-
c	41	14.4	84.7	338	79	AW283100	AW283100 LG1_225_D
	42	14.4	84.7	346	62	AI920823	AI920823 wn82f01.x
	43	14.4	84.7	348	35	C25936	C25936 C25936 Rice
	44	14.4	84.7	348	44	AI247014	AI247014 qx52c08.x
	45	14.4	84.7	353	47	AI539116	AI539116 tp78c02.x

ALIGNMENTS

```
RESULT   1
AI951640
LOCUS       AI951640      575 bp    mRNA            EST       27-OCT-1999
DEFINITION  wv37f09.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:2531753 3'
            similar to contains element TAR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI951640
VERSION     AI951640.1  GI:5743950
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 19, 1999 this sequence version replaced gi:4458780.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2531753"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT       75 a    190 c    246 g     63 t      1 others
ORIGIN


  Query Match           100.0%;  Score 17;  DB 63;  Length 575;
  Best Local Similarity  100.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||||||||||
Db     28 GGGGCCGTCCTGGTGCC 44


RESULT   2
AW124365
LOCUS       AW124365      275 bp    mRNA            EST       22-OCT-1999
DEFINITION  UI-M-BH2.1-apq-h-12-0-UI.s1 NIH_BMAP_M_S3.1 Mus musculus cDNA clone
            UI-M-BH2.1-apq-h-12-0-UI 3', mRNA sequence.
ACCESSION   AW124365
VERSION     AW124365.1  GI:6099895
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5866533.
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
```

```
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .275
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-BH2.1-apq-h-12-0-UI"
                      /clone_lib="NIH_BMAP_M_S3.1"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_M_S3.1 library is a subtracted library of a
                      series, ultimately derived from a mixture of individually
                      tagged normalized libraries from ten regions of the mouse
                      brain (cerebellum, brain stems, olfactory bulbs,
                      hypothalamus, cortex, amygdala, basal ganglia, pineal
                      gland, striatum, hipoccampus) after a series of
                      subtractions to reduce the representation of cDNAs from
                      which ESTs had already been generated. The following
                      serially subtracted libraries were generated in this
                      process: NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                      The subtracted library (NIH_BMAP_M_S3.1) was constructed
                      as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                      clones from which 3' ESTs had been derived was used as a
                      driver in a hybridization with the NIH_BMAP_M_S2 library
                      in the form of single-stranded circles. The remaining
                      single-stranded circles (subtracted library) was purified
                      by hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (LifeTechnologies) to generate the
                      NIH_BMAP_M_S3.1 library. This procedure has been
                      previously described (Bonaldo, Lennon and Soares, Genome
                      Research 6:791-806, 1996)
                      TAG_LIB=NIH_BMAP_M_S3.1
                      TAG_TISSUE=brain-stems
                      TAG_SEQ=TCATG"
BASE COUNT       41 a     92 c     94 g     47 t      1 others
ORIGIN


  Query Match             94.1%;  Score 16;  DB 69;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGC 16
          ||||||||||||||||
Db     52 GGGGCCGTCCTGGTGC 67


RESULT   3
AA911158/c
LOCUS       AA911158      246 bp    mRNA            EST       21-APR-1998
DEFINITION  oe69e07.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1416900 3'
            similar to gb:L19686_rna1 MACROPHAGE MIGRATION INHIBITORY FACTOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA911158
VERSION     AA911158.1  GI:3050448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 246)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2282319.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 337   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .246
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1416900"
                      /clone_lib="NCI_CGAP_Lu5"
                      /tissue_type="carcinoid"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      neuroendocrine lung carcinoid, and was then primed with a
                      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                      to Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector. Library is normalized. Library was
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       40 a     86 c     72 g     48 t
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 40;  Length 246;
  Best Local Similarity   94.1%;  Pred. No. 9.1e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          ||||| ||||||||||
Db    137 GGGGCTGTCCTGGTGCC 121


RESULT   4
AI836163
LOCUS       AI836163      293 bp    mRNA            EST       14-JUL-1999
DEFINITION  UI-M-AQ0-aai-g-01-0-UI.s2 NIH_BMAP_MHI Mus musculus cDNA clone
            UI-M-AQ0-aai-g-01-0-UI 3', mRNA sequence.
ACCESSION   AI836163
VERSION     AI836163.1  GI:5470376
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3186956.
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
```

```
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            The following repetitive elements were found in this cDNA sequence:
            224-268, >(TGG)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .293
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-AQ0-aai-g-01-0-UI"
                      /clone_lib="NIH_BMAP_MHI"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_MHI library is a non-normalized library
                      constructed from  mouse hippocampus. The tag is a string
                      of 5 nucleotides present  between the Not I site and the
                      oligo-dT track.  The library was  constructed as described
                      by Bonaldo, Lennon and Soares,  Genome Research 6:
                      791-806, 1996. Tissue provided by Ms. Annie Novakovich,
                      Zivic-Miller Laboratories.
                      TAG_LIB=NIH_BMAP_MHI
                      TAG_TISSUE=olfactory-bulbs
                      TAG_SEQ=CATGG"
BASE COUNT       42 a     69 c    128 g     54 t
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 61;  Length 293;
  Best Local Similarity   94.1%;  Pred. No. 9.6e+02;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           ||||||||||| |||||
Db     276 GGGGCCGTCCTCGTGCC 292


RESULT   5
AQ163728/c
LOCUS       AQ163728      333 bp    DNA             GSS       16-OCT-1998
DEFINITION  HS_2247_B1_G12_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2247 Col=23 Row=N, genomic survey
            sequence.
ACCESSION   AQ163728
VERSION     AQ163728.1  GI:3561103
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 333)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2247  row: N  column: 23
            Class: BAC ends
            High quality sequence stop: 333.
FEATURES              Location/Qualifiers
     source           1. .333
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=2247 Col=23 Row=N"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT       86 a     85 c     85 g     76 t      1 others
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 102;  Length 333;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           ||||| ||||||||||
Db      97 GGGGCAGTCCTGGTGCC 81


RESULT   6
AI377512
LOCUS       AI377512      388 bp    mRNA            EST       18-MAR-1999
DEFINITION  tc37d11.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:2066805 3', mRNA sequence.
ACCESSION   AI377512
VERSION     AI377512.1  GI:4187365
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 15, 1998 this sequence version replaced gi:3226956.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 915  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 268.
FEATURES              Location/Qualifiers
     source           1. .388
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2066805"
                      /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                      /dev_stage="8-9 weeks"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from pooled 8-9 week
                      (total) fetus material with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']."
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       37 a    149 c    155 g     47 t
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 45;  Length 388;
  Best Local Similarity   94.1%;  Pred. No. 1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      1 GGGGCCGTCCTGGTGCC 17
          ||||||||||||| |||
Db     28 GGGGCCGTCCTGGAGCC 44


RESULT   7
AQ143042/c
LOCUS       AQ143042      430 bp    DNA             GSS       24-SEP-1998
DEFINITION  HS_3064_A1_A07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3064 Col=13 Row=A, genomic survey
            sequence.
ACCESSION   AQ143042
VERSION     AQ143042.1  GI:3533695
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K , Swartzell,S., Holzman T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3064  row: A  column: 13
            Class: BAC ends
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3064 Col=13 Row=A"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       70 a    161 c     69 g    122 t      8 others
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 102;  Length 430;
  Best Local Similarity   94.1%;  Pred. No. 1.1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          ||||| ||||||||||
Db     27 GGGGCAGTCCTGGTGCC 11


RESULT   8
AW260226/c
LOCUS       AW260226      476 bp    mRNA            EST       23-DEC-1999
DEFINITION  um75a09.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2301016 3', mRNA sequence.
ACCESSION   AW260226
VERSION     AW260226.1  GI:6633207
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 9, 1999 this sequence version replaced gi:5433886.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1009788
            Seq primer: custom primer used
            High quality sequence stop: 373.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2301016"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT       89 a    118 c    107 g    162 t
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 79;  Length 476;
  Best Local Similarity   94.1%;  Pred. No. 1.1e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||| ||||||||||
Db    160 GGGGCCCTCCTGGTGCC 144


RESULT   9
AA544009
LOCUS       AA544009      630 bp    mRNA            EST       01-AUG-1997
DEFINITION  vj83h11.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:935685 5', mRNA sequence.
ACCESSION   AA544009
VERSION     AA544009.1  GI:2292486
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The W shU-HHMI Mouse EST Project
```

```
  JOURNAL   Unpublished (1996)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:636298.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:540605
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 216.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:935685"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      137 a    164 c    167 g    162 t
ORIGIN


  Query Match             90.6%;  Score 15.4;  DB 35;  Length 630;
  Best Local Similarity   94.1%;  Pred. No. 1.2e+03;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGCC 17
           ||||| ||||||||||
Db     549 GGGGCAGTCCTGGTGCC 565


RESULT  10
AA518462
LOCUS       AA518462      295 bp    mRNA            EST       16-JUL-1997
DEFINITION  vi06h11.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:903045 5' similar to SW:AQP5_RAT P47864 AQUAPORIN 5. ;, mRNA
            sequence.
ACCESSION   AA518462
VERSION     AA518462.1  GI:2259147
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1315404.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:523709
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:903045"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins. "
BASE COUNT       5 a     84 c     82 g     75 t
ORIGIN


  Query Match             88.2%;  Score 15;  DB 34;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTG 15
           |||||||||||||||
Db     272 GGGGCCGTCCTGGTG 286


RESULT  11
AQ911288/c
LOCUS       AQ911288      401 bp    DNA             GSS       30-NOV-1999
DEFINITION  GSSTc03322 Trypanosome cruzi random genomic library Trypanosoma
            cruzi genomic clone G47H14 5', genomic survey sequence.
ACCESSION   AQ911288
VERSION     AQ911288.1  GI:6491618
KEYWORDS    GSS.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Sanchez,D.O.
  TITLE     Trypanosoma cruzi random genomic sequences
  JOURNAL   Unpublished (1999)
COMMENT     On Feb 19, 1999 this sequence version replaced gi:4129449.
            Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            CP(1650) San Martin, Prov. de BS AS. Argentina
            Tel: (54-11)4752-0021
            Fax: (54-11)4752-9639
            Email: dsanchez@inti.gov.ar
            Seq primer: T7
            Class: shotgun.
FEATURES             Location/Qualifiers
```

```
     source          1. .401
                     /organism="Trypanosoma cruzi"
                     /strain="Cl-Brener"
                     /db_xref="taxon:5693"
                     /clone="G47H14"
                     /clone_lib="Trypanosome cruzi random genomic library"
                     /cell_type="epimastigote"
                     /note="Vector: pBS(-) (Stratagene)"
BASE COUNT       61 a    129 c    131 g     57 t     23 others
ORIGIN


  Query Match             88.2%;  Score 15;  DB 92;  Length 401;
  Best Local Similarity   88.2%;  Pred. No. 1.6e+03;
  Matches   15;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTGCC 17
          |||||||||: ||||||
Db     46 GGGGCCGTCMCGGTGCC 30


RESULT  12
AI004179/c
LOCUS       AI004179      483 bp    mRNA            EST       27-AUG-1998
DEFINITION  ou56c12.x1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1631830 3',
            mRNA sequence.
ACCESSION   AI004179
VERSION     AI004179.1  GI:3213689
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397624.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 815   Std Error: 0.00
            Seq primer: primer name ambiguous
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1631830"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo. "
BASE COUNT      110 a    132 c    162 g     79 t
ORIGIN


  Query Match             88.2%;  Score 15;  DB 41;  Length 483;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTG 15
          |||||||||||||||
Db    452 GGGGCCGTCCTGGTG 438


RESULT  13
AW328031
LOCUS       AW328031      536 bp    mRNA            EST       28-JAN-2000
DEFINITION  dr03g02.x1 NIH_MGC_3 Homo sapiens cDNA clone IMAGE:2847170 5', mRNA
            sequence.
ACCESSION   AW328031
VERSION     AW328031.1  GI:6798527
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137356.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Edge BioSystems
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center (NISC)
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Plate: LLCM0029  row: N  column: 3
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2847170"
                     /clone_lib="NIH_MGC_3"
                     /tissue_type="Burkitt lymphoma"
                     /cell_line="MGC4"
                     /note="Organ: Lymph; Vector: pOTB7a; Library prepared by
                     Edge BioSystems."
BASE COUNT       85 a    205 c    137 g    108 t      1 others
ORIGIN


  Query Match             88.2%;  Score 15;  DB 74;  Length 536;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGGCCGTCCTGGTG 15
          |||||||||||||||
Db    234 GGGGCCGTCCTGGTG 248


RESULT  14
AL041617
LOCUS       AL041617      636 bp    mRNA            EST       29-SEP-1999
DEFINITION  DKFZp434K1717_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434K1717 5', mRNA sequence.
ACCESSION   AL041617
VERSION     AL041617.1  GI:5420966
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Bloecker, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133694.
            Contact: Bloecker H
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by GBF within the cDNA sequencing consortium of the
            German Genome Project.
            s1 sequence also available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES              Location/Qualifiers
     source           1. .636
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="DKFZp434K1717"
                      /clone_lib="434 (synonym: htes3)"
                      /tissue_type="testis"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT       100 a    238 c    163 g    135 t
ORIGIN


  Query Match             88.2%;  Score 15;  DB 61;  Length 636;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+03;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTG 15
           |||||||||||||||
Db     341 GGGGCCGTCCTGGTG 355


RESULT  15
AI950970
LOCUS       AI950970      106 bp    mRNA            EST       06-SEP-1999
DEFINITION  wx61f11.x1 NCI_CGAP_Br18 Homo sapiens cDNA clone IMAGE:2548173 3'
            similar to gb:M63193 THYMIDINE PHOSPHORYLASE (HUMAN);, mRNA
            sequence.
ACCESSION   AI950970
VERSION     AI950970.1  GI:5743280
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/nci gap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948680.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .106
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2548173"
                      /clone_lib="NCI_CGAP_Br18"
                      /tissue_type="four pooled high-grade tumors, including two
                      primary tumors and two metastatic to ovary"
                      /lab_host="DH10B"
                      /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                       Library constructed by Life Technologies."
BASE COUNT       15 a     32 c     48 g     11 t
ORIGIN


  Query Match             84.7%;  Score 14.4;  DB 63;  Length 106;
  Best Local Similarity   93.8%;  Pred. No. 2e+03;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGCCGTCCTGGTGC 16
           |||||||||| |||||
Db      84 GGGGCCGTCCCGGTGC 99


Search completed: June 22, 2000, 12:08:30
Job time: 10424 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 14:58:35 ; Search time 8627.09 Seconds
(without alignments)
-2.368 Million cell updates/sec

Title: US-09-362-485-17
Perfect score: 21
Sequence: 1 GACGTCGACCTACGCGCTGAC 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	21	5	A87620	A87620 Sequence 17
2	21	100.0	1194	2	MTU92472	U92472 Mycobacteri
3	21	100.0	1208	5	A89752	A89752 Sequence 9
4	21	100.0	1209	5	A87612	A87612 Sequence 9
5	21	100.0	1228	5	A87608	A87608 Sequence 5
6	21	100.0	1228	5	A89748	A89748 Sequence 5
7	21	100.0	1229	5	A87610	A87610 Sequence 7
8	21	100.0	1229	5	A89750	A89750 Sequence 7
9	21	100.0	1235	5	A87606	A87606 Sequence 3
10	21	100.0	1235	5	A87609	A87609 Sequence 6
11	21	100.0	1235	5	A87611	A87611 Sequence 8
12	21	100.0	1235	5	A89746	A89746 Sequence 3
13	21	100.0	1235	5	A89749	A89749 Sequence 6
14	21	100.0	1235	5	A89751	A89751 Sequence 8
15	21	100.0	1236	5	A87613	A87613 Sequence 10
16	21	100.0	1236	5	A89753	A89753 Sequence 10
17	21	100.0	1237	5	A87607	A87607 Sequence 4
18	21	100.0	1237	5	A89747	A89747 Sequence 4
19	21	100.0	1245	5	A87605	A87605 Sequence 2
20	21	100.0	1245	5	A89745	A89745 Sequence 2
21	21	100.0	1260	5	A87604	A87604 Sequence 1
22	21	100.0	1260	5	A89744	A89744 Sequence 1
23	21	100.0	2584	1	MTALADH	X63069 M.tuberculo
24	21	100.0	56414	1	MTV002	AL008967 Mycobacte
25	17.4	82.9	30590	1	SC4G2	AL031371 Streptomy
c 26	16.8	80.0	3606	5	E03846	E03846 Fusaric aci
c 27	16.8	80.0	3608	1	KPNFDT1	D84271 Klebsiella
c 28	16.8	80.0	35121	34	CEF10G8	Z80216 Caenorhabdi
c 29	16.8	80.0	40745	1	SCI51	AL109848 Streptomy
c 30	16.8	80.0	110000	32	CEY106G6_1	Continuation (2 of
c 31	16.2	77.1	1880	8	AF171874	AF171874 Zea mays
c 32	16.2	77.1	2084	34	AB017021	AB017021 Ephydatia
c 33	16.2	77.1	2538	24	E09053	E09053 gDNA encodi
34	16.2	77.1	3105	1	HALHMA	M87833 Halobacteri
c 35	16.2	77.1	4233	7	ZMU16123	U16123 Zea mays in
c 36	16.2	77.1	6358	1	LLYK90C	X56660 L.lactamgen
37	16.2	77.1	8046	2	AF049107	AF049107 Myxococcu
c 38	16.2	77.1	8931	2	AF000579	AF000579 Bordetell
c 39	16.2	77.1	12065	1	AB015802	AB015802 Acetobact
40	16.2	77.1	13149	2	AE002086	AE002086 Deinococc
c 41	16.2	77.1	23730	1	SC3F7	AL021409 Streptomy
42	16.2	77.1	24771	42	AC013014	AC013014 Drosophil
43	16.2	77.1	25306	2	AF058302	AF058302 Streptomy
c 44	16.2	77.1	36890	42	AC014994	AC014994 Drosophil
45	16.2	77.1	40875	1	SCE65	AL133236 Streptomy

ALIGNMENTS

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NEFRV 5
          |||||
Db     14 NEFRV 18
```

RESULT 7
B71624
hypothetical protein PFB0080c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C;Accession: B71624
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: B71624
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <GAR>
A;Cross-references: GB:AE001369; GB:AE001362; NID:g3845084; PID:g3845088; TIGR:PFB0080c
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0080c

Query Match 100.0%; Score 26; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NEFRV 5
          |||||
Db    154 NEFRV 158
```

RESULT 8
A43830
alanine dehydrogenase (EC 1.4.1.1) ald - Mycobacterium tuberculosis
C;Species: Mycobacterium tuberculosis
C;Date: 30-Jun-1993 #sequence_revision 24-Jul-1998 #text_change 11-Jun-1999
C;Accession: C70883; A43830; B43830; S18864
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: C70883
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-371 <COL>
A;Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAA15575.1; PID:e117391
A;Experimental source: strain H37Rv
R;Andersen, A.B.; Andersen, P.; Ljungqvist, L.
Infect. Immun. 60, 2317-2323, 1992
A;Title: Structure and function of a 40,000-molecular-weight protein antigen of Mycobact
A;Reference number: A43830; MUID:92267644
A;Accession: A43830
A;Molecule type: DNA
A;Residues: 1-14,'QF',15-371 <AND>
A;Cross-references: EMBL:X63069; NID:g44565; PIDN:CAA44791.1; PID:g44566
A;Accession: B43830
A;Molecule type: protein
A;Residues: 1,'X',3-5,'X',7,'XX',10 <AND2>
C;Comment: This enzyme may be important in synthesis of peptidoglycan in the cell wall.
C;Genetics:
A;Gene: ald
C;Superfamily: alanine dehydrogenase; alanine dehydrogenase homology
C;Keywords: NAD; oxidoreductase
F;1-278/Domain: alanine dehydrogenase homology <ALA>
F;170-198/Region: beta-alpha-beta NAD nucleotide-binding fold

Query Match 100.0%; Score 26; DB 1; Length 371;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NEFRV 5
          |||||
Db     12 NEFRV 16
```

RESULT 9
T36876
probable L-alanine dehydrogenase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36876
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21617
A;Accession: T36876
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-371 <MUR>
A;Cross-references: EMBL:AL109848; PIDN:CAB52837.1; GSPDB:GN00070; SCOEDB:SCI51.13c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCI51.13c

Query Match 100.0%; Score 26; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NEFRV 5
          |||||
Db     12 NEFRV 16
```

RESULT 10
PQ0223
polyprotein - simian hepatitis A virus (fragments)
N;Contains: VP0 protein; VP1 protein; VP3 protein
C;Species: simian hepatitis A virus
A;Note: host Macaca fascicularis (crab-eating macaque)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 07-Feb-1997
C;Accession: PQ0223
R;Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.
J. Gen. Virol. 72, 1685-1689, 1991
A;Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus
A;Reference number: JQ1180; MUID:91311421
A;Accession: PQ0223
A;Molecule type: genomic RNA
A;Residues: 1-384 <NAI>
C;Superfamily: hepatitis A virus genome polyprotein
C;Keywords: polyprotein
F;1-76/Product: VP0 protein (fragment) #status predicted <VP0>
F;77-169/Product: VP3 protein (fragment) #status predicted <VP3>
F;170-384/Product: VP1 protein (fragment) #status predicted <VP1>

Query Match 100.0%; Score 26; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NEFRV 5
          |||||
Db     80 NEFRV 84
```

RESULT 11
T15756

hypothetical protein C34D10.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15756
R;Minx, P.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid C34D10.
A;Reference number: Z18398
A;Accession: T15756
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-720 <MIN>
A;Cross-references: EMBL:U29157; NID:g861409; PID:g861411; PIDN:AAA68423.1; CESP:C34D10.
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C34D10.2
A;Introns: 26/2; 64/2; 96/2; 133/3; 198/1; 243/2; 294/3; 339/1; 398/3; 488/1; 526/3; 648

Query Match 100.0%; Score 26; DB 2; Length 720;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 29 NEFRV 33

RESULT 12
GNNYS2
genome polyprotein - simian hepatitis A virus (strain CY-145) (fragment)
N;Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein
C;Species: simian hepatitis A virus
A;Note: host Macaca fascicularis (cynomolgus macaque)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999
C;Accession: JQ1180
R;Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.
J. Gen. Virol. 72, 1685-1689, 1991
A;Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus mac
A;Reference number: JQ1180; MUID:91311421
A;Accession: JQ1180
A;Molecule type: genomic RNA
A;Residues: 1-839 <NAI>
A;Cross-references: GB:M59286; NID:g329599; PIDN:AAA45473.1; PID:g555083
C;Superfamily: hepatitis A virus genome polyprotein
C;Keywords: coat protein; core protein; glycoprotein; polyprotein
F;1-23/Product: coat protein 1A #status predicted <VP0>
F;24-245/Product: coat protein 1B #status predicted <VP3>
F;246-491/Product: coat protein 1C #status predicted <VP1>
F;492-839/Product: core protein 2A (fragment) #status predicted <P2P>
F;261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 26; DB 1; Length 839;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 249 NEFRV 253

RESULT 13
GNNYHA
genome polyprotein - human hepatitis A virus (strain CR326) (fragment)
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A
C;Species: human hepatitis A virus
A;Note: host Homo sapiens (man)
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C;Accession: A03904
R;Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.W
J. Virol. 54, 247-255, 1985
A;Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.
A;Reference number: A03904; MUID:85185648
A;Accession: A03904
A;Molecule type: genomic RNA
A;Residues: 1-852 <LIN>
A;Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593
C;Superfamily: hepatitis A virus genome polyprotein
C;Keywords: coat protein; core protein; polyprotein
F;1-245/Product: coat protein 1A #status predicted <C1A>
F;246-491/Product: coat protein 1B #status predicted <C1B>
F;492-836/Product: coat protein 1C #status predicted <C1C>
F;837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 26; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 249 NEFRV 253

RESULT 14
S63064
probable membrane protein YNL123w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N1897
C;Species: Saccharomyces cerevisiae
C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 12-Dec-1997
C;Accession: S63064; S67335
R;De Antoni, A.; D Angelo, M.; Dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63047
A;Accession: S63064
A;Molecule type: DNA
A;Residues: 1-997 <DEA>
A;Cross-references: EMBL:Z71399; NID:g1302053; PID:e239802; PID:g1302054; MIPS:YNL123
A;Experimental source: strain S288C
R;d'Antoni, A.; d'Angelo, M.; dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi,
submitted to the EMBL Data Library, February 1996
A;Description: The DNA sequence of cosmid 14-13b from chromosome XIV of Saccharomyces
A;Reference number: S67327
A;Accession: S67335
A;Molecule type: DNA
A;Residues: 1-997 <DAN>
A;Cross-references: EMBL:Z69382; NID:g1183941; PID:e221825; PID:g1183950
C;Genetics:
A;Map position: 14L
A;Note: YNL123w
C;Keywords: transmembrane protein
F;598-614/Domain: transmembrane #status predicted <TMM>

Query Match 100.0%; Score 26; DB 2; Length 997;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 498 NEFRV 502

RESULT 15
T31688
Ca2+-transporting ATPase (EC 3.6.1.38), plasma membrane - Paramecium tetraurelia (SGC
C;Species: Paramecium tetraurelia
C;Date: 01-Nov-1999 #sequence_revision 01-Nov-1999 #text_change 01-Nov-1999
C;Accession: T31688
R;Elwess, N.L.; van Houten, J.L.
Journal of Eukaryotic Microbiology 44, 250-257, 1997
A;Title: Cloning and molecular analysis of the plasma membrane Ca(2+)-ATPase gene in
A;Reference number: Z21062
A;Accession: T31688
A;Status: preliminary

A;Residues: 1-1160 <ELW>
A;Cross-references: EMBL:U05880; NID:g505622; PID:g505623; PIDN:AAB81284.1

Query Match 100.0%; Score 26; DB 3; Length 1160;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |||||
Db    1007 NEFRV 1011
```

Search completed: June 22, 2000, 19:28:58
Job time: 5201 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:33:55 ; Search time 40.62 Seconds
(without alignments)
3.749 Million cell updates/sec

Title: US-09-362-485-29
Perfect score: 26
Sequence: 1 NEFRV 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	26	100.0	172	1	RL10_CGDAB	P36249	citrus gree
2	26	100.0	267	1	AVP3_PSESH	Q52430	pseudomonas
3	26	100.0	371	1	DHA_MYCTU	P30234	mycobacteri
4	26	100.0	808	1	POLG_HPAVG	Q02381	hepatitis a
5	26	100.0	839	1	POLG_HPAVT	P31788	simian hepa
6	26	100.0	852	1	POLG_HPAVC	P06442	hepatitis a
7	26	100.0	997	1	YNM3_YEAST	P53920	saccharomyc
8	26	100.0	1046	1	POL_SIVAG	P27980	simian immu
9	26	100.0	2226	1	POLG_HPAV2	P26580	hepatitis a
10	26	100.0	2226	1	POLG_HPAV4	P26581	hepatitis a
11	26	100.0	2226	1	POLG_HPAV8	P26582	hepatitis a
12	26	100.0	2227	1	POLG_HPAVH	P08617	hepatitis a
13	26	100.0	2227	1	POLG_HPAVL	P06441	hepatitis a
14	26	100.0	2227	1	POLG_HPAVM	P13901	hepatitis a
15	26	100.0	2230	1	POLG_HPAVS	P14553	simian hepa
16	25	96.2	309	1	IKI1_YEAST	P38874	saccharomyc
17	25	96.2	573	1	YKK3_CAEEL	P34280	caenorhabdi
18	25	96.2	605	1	VCLA_GOSHI	P09799	gossypium h
19	25	96.2	884	1	TERT_YEAST	Q06163	saccharomyc
20	23	88.5	211	1	END3_HAEIN	P44319	haemophilus
21	23	88.5	226	1	ATP6_APILI	Q00275	apis mellif
22	23	88.5	226	1	COX2_LASSP	P29875	lasius sp.
23	23	88.5	228	1	COX2_YPOMA	P98048	yponomeuta
24	23	88.5	229	1	COX2_ONCFA	P29876	oncopeltus
25	23	88.5	246	1	PYHD_NPVMB	P12423	mamestra br
26	23	88.5	246	1	PYHD_NPVPF	P22115	panolis fla
27	23	88.5	291	1	HMP1_BOVIN	P10036	bos taurus
28	23	88.5	291	1	HMP1_HUMAN	P28069	homo sapien
29	23	88.5	291	1	HMP1_MOUSE	Q00286	mus musculu
30	23	88.5	291	1	HMP1_SHEEP	P79364	ovis aries
31	23	88.5	294	1	BLAN_SERMA	P52682	serratia ma
32	23	88.5	317	1	HMP1_RAT	P10037	rattus norv
33	23	88.5	324	1	MAT1_MARGL	P51950	marthasteri
34	23	88.5	329	1	AN36_HELPY	P94851	helicobacte
35	23	88.5	353	1	CEBA_BOVIN	O02754	bos taurus
36	23	88.5	357	1	YAOC_SCHPO	Q10091	schizosacch
37	23	88.5	358	1	CEBA_HUMAN	P49715	homo sapien
38	23	88.5	358	1	CEBA_RAT	P05554	rattus norv
39	23	88.5	359	1	CEBA_MOUSE	P53566	mus musculu
40	23	88.5	383	1	NTRB_RHILP	P41503	rhizobium l
41	23	88.5	384	1	YZ43_METJA	Q60298	methanococc
42	23	88.5	394	1	GBAS_BOVIN	P04896	bos taurus
43	23	88.5	394	1	GBAS_CRILO	P16052	cricetulus
44	23	88.5	394	1	GBAS_HUMAN	P04895	homo sapien
45	23	88.5	394	1	GBAS_MOUSE	P04894	mus musculu

ALIGNMENTS

RESULT 1
RL10_CGDAB
ID RL10_CGDAB STANDARD; PRT; 172 AA.
AC P36249;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L10.
GN RPLJ.
OS Citrus greening disease-associated bacterium-like organism.
OC Bacteria; Proteobacteria; gamma subdivision.
RN [1]
RP SEQUENCE FROM N.A.
RA Villechanoux S., Garnier M., Laigret F., Renaudin J., Bove J.M.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M94319; AAA23107.1; -.
DR PROSITE; PS01109; RIBOSOMAL_L10; 1.
DR PFAM; PF00466; Ribosomal_L10; 1.
KW Ribosomal protein.
SQ SEQUENCE 172 AA; 18475 MW; 7BC9F2D229E359CC CRC64;

Query Match 100.0%; Score 26; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 101 NEFRV 105

RESULT 2
AVP3_PSESH
ID AVP3_PSESH STANDARD; PRT; 267 AA.
AC Q52430;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AVIRULENCE PROTEIN AVRPPH3.
GN AVRPPH3.
OS Pseudomonas syringae (pv. phaseolicola).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RACE 3 ISOLATE 1301A, AND RACE 4 ISOLATE 1302A;
RX MEDLINE; 92208315.

```
RA   Jenner C., Hitchin E., Mansfield J., Walters K., Betteridge P.,
RA   Teverson D., Taylor J.;
RT   "Gene-for-gene interactions between Pseudomonas syringae pv.
RT   phaseolicola and Phaseolus.";
RL   Mol. Plant Microbe Interact. 4:553-562(1991).
CC   -!- FUNCTION: RESPONSIBLE FOR THE CULTIVAR-SPECIFIC AVIRULENCE TO
CC       PHASEOLUS VULGARIS CV. TENDERGREEN.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86401; AAA25727.1; -.
KW   Virulence.
SQ   SEQUENCE   267 AA;  28703 MW;  A503220A905C60A8 CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     262 NEFRV 266


RESULT   3
DHA_MYCTU
ID   DHA_MYCTU      STANDARD;      PRT;   371 AA.
AC   P30234; O33322;
DT   01-APR-1993 (Rel. 25, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ALANINE DEHYDROGENASE (EC 1.4.1.1) (40 KD ANTIGEN).
GN   ALD OR RV2780 OR MTV002.45.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ERDMANN;
RX   MEDLINE; 92267644.
RA   Andersen A.B., Andersen P., Ljungqvist L.;
RT   "Structure and function of a 40,000-molecular-weight protein antigen
RT   of Mycobacterium tuberculosis.";
RL   Infect. Immun. 60:2317-2323(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Singh M., Hutter B.;
RT   "Host-vector system for high level expression and purification of
RT   enzymatically active L-alanine dehydrogenase of M.tuberculosis in
RT   E.coli.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN CELL WALL SYNTHESIS AS L-ALANINE IS
CC       AN IMPORTANT CONSTITUENT OF THE PEPTIDOGLYCAN LAYER.
CC   -!- CATALYTIC ACTIVITY: L-ALANINE + H(2)O + NAD(+) = PYRUVATE +
CC       NH(3) + NADH.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: STRONG, TO OTHER ALANINE DEHYDROGENASES AND PARTIAL
CC       WITH PYRIDINE NUCLEOTIDE TRANSHYDROGENASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63069; CAA44791.1; -.
DR   EMBL; U92472; AAC38804.1; -.
DR   EMBL; AL008967; CAA15575.1; -.
DR   PIR; A43830; A43830.
DR   PIR; S18864; S18864.
DR   PROSITE; PS00836; ALADH_PNT_1; 1.
DR   PROSITE; PS00837; ALADH_PNT_2; 1.
DR   PFAM; PF01262; AlaDh_PNT; 1.
DR   TUBERCULIST; Rv2780; -.
KW   Oxidoreductase; NAD; Antigen.
FT   ACT_SITE     96     96       POTENTIAL.
FT   NP_BIND     170    200       NAD (BY SIMILARITY).
FT   CONFLICT     13     13       E -> EFQ (IN REF. 1).
SQ   SEQUENCE   371 AA;  38713 MW;  9DF7540524DC116A CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 371;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db      12 NEFRV 16


RESULT   4
POLG_HPAVG
ID   POLG_HPAVG     STANDARD;      PRT;   808 AA.
AC   Q02381;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE   P2A] (FRAGMENT).
OS   Hepatitis A virus (strain GA76).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92260183.
RA   Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;
RT   "Characterization of a genetic variant of human hepatitis A virus.";
RL   J. Med. Virol. 36:118-124(1992).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; M66695; AAA45477.1; -.
KW   Polyprotein; Coat protein; Core protein.
FT   NON_TER       1      1
FT   CHAIN        <1      2       COAT PROTEIN VP4 (P1A).
FT   CHAIN         3    223       COAT PROTEIN VP2 (P1B).
FT   CHAIN       224    470       COAT PROTEIN VP3 (P1C).
FT   CHAIN       471    770       COAT PROTEIN VP1 (P1D).
FT   CHAIN       771   >808       CORE PROTEIN P2A.
FT   NON_TER     808    808
SQ   SEQUENCE   808 AA;  90632 MW;  D80CE7E57A479C12 CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 808;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NEFRV 5
          |||||
Db    228 NEFRV 232


RESULT   5
POLG_HPAVT
ID   POLG_HPAVT     STANDARD;      PRT;   839 AA.
AC   P31788;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE   P2A] (FRAGMENT).
OS   Simian hepatitis A virus (strain CY-145).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91311421.
RA   Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;
RT   "Sequence analysis of a new hepatitis A virus naturally infecting
RT   cynomolgus macaques (Macaca fascicularis).";
RL   J. Gen. Virol. 72:1685-1689(1991).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59286; AAA45473.1; -.
DR   PIR; JQ1180; GNNYS2.
KW   Polyprotein; Coat protein; Core protein.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492      ?       COAT PROTEIN VP1 (P1D).
FT   CHAIN         ?   >839       CORE PROTEIN P2A.
FT   NON_TER     839    839
SQ   SEQUENCE   839 AA;  93825 MW;  2CACC4BD1E192DBC CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 839;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NEFRV 5
          |||||
Db    249 NEFRV 253


RESULT   6
POLG_HPAVC
ID   POLG_HPAVC     STANDARD;      PRT;   852 AA.
AC   P06442; Q83741; Q83742;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE   P2A] (FRAGMENT).
OS   Hepatitis A virus (strain CR326).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85185648.
RA   Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,
RA   Young A., Mitra S.W.;
RT   "Molecular cloning and partial sequencing of hepatitis A viral cDNA.";
RL   J. Virol. 54:247-255(1985).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M1003 ; AAA45470.1; -.
DR   PIR; A03904  GNNYHA.
KW   Polyprotein, Coat protein; Core protein.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    836       COAT PROTEIN VP1 (P1D).
FT   CHAIN       837   >852       CORE PROTEIN P2A.
FT   NON_TER     852    852
SQ   SEQUENCE   852 AA;  95563 MW;  73D3ED0AD532820E CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 852;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NEFRV 5
          |||||
Db    249 NEFRV 253


RESULT   7
YNM3_YEAST
ID   YNM3_YEAST     STANDARD;      PRT;   997 AA.
AC   P53920;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 110.9 KD PROTEIN IN SPC98-TOM70 INTERGENIC REGION.
GN   YNL123W OR N1897.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97245296.
RA   de Antoni A., D Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA   Pallavicini A., Lanfranchi G., Valle G.;
RT   "The DNA sequence of cosmid 14-13b from chromosome XIV of
```

```
RT   Saccharomyces cerevisiae reveals an unusually high number of
RT   overlapping open reading frames.";
RL   Yeast 13:261-266(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69382; CAA93384.1; -.
DR   EMBL; Z71399; CAA96004.1; -.
DR   PFAM; PF00595; PDZ; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   997 AA;  110881 MW;  A26005C1DDDB932C CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 997;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     498 NEFRV 502


RESULT   8
POL_SIVAG
ID   POL_SIVAG      STANDARD;      PRT;  1046 AA.
AC   P27980;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);
DE   REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].
GN   POL.
OS   Simian immunodeficiency virus (AGM3 isolate) (SIV-AGM).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90232731.
RA   Baier M., Garber C., Mueller C., Cichutek K., Kurth R.;
RT   "Complete nucleotide sequence of a simian immunodeficiency virus from
RT   African green monkeys: a novel type of intragroup divergence.";
RL   Virology 176:216-221(1990).
CC   -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
CC       DETERMINED.
CC   -!- MISCELLANEOUS: THIS IS AN AFRICAN GREEN MONKEY ISOLATE.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC       KNOWN AS THE RETROPEPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M30931; AAA91914.1; -.
DR   HSSP; P03366; 1HMV.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PFAM; PF00075; rnaseH; 1.
DR   PFAM; PF00077; rvp; 1.
DR   PFAM; PF00078; rvt; 1.
DR   PFAM; PF00552; integrase; 1.
DR   PFAM; PF00665; rve; 1.
KW   AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;
KW   Nuclease; Transferase; RNA-directed DNA polymerase.
FT   CHAIN         1    195       PROTEASE.
FT   ACT_SITE    118    118       BY SIMILARITY.
SQ   SEQUENCE   1046 AA;  119328 MW;  9068415C43AD0A6B CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 1046;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     558 NEFRV 562


RESULT   9
POLG_HPAV2
ID   POLG_HPAV2     STANDARD;      PRT;  2226 AA.
AC   P26580;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain 24a).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91162758.
RA   Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA   Cromeans T., Jansen R.W.;
RT   "Antigenic and genetic variation in cytopathic hepatitis A virus
RT   variants arising during persistent infection: evidence for genetic
RT   recombination.";
RL   J. Virol. 65:2056-2065(1991).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59810; AAA45468.1; -.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    794       COAT PROTEIN VP1 (P1D).
FT   CHAIN       795    900       CORE PROTEIN P2A.
FT   CHAIN       901   1087       CORE PROTEIN P2B.
FT   CHAIN      1088   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1495       PROBABLE PROTEIN P3A.
FT   CHAIN      1496   1518       PROBABLE PROTEIN P3B.
FT   CHAIN      1519   1737       PROBABLE PROTEIN P3C.
FT   CHAIN      1738   2226       RNA-DIRECTED POLYMERASE P3D.
SQ   SEQUENCE   2226 AA;  251152 MW;  6CD85A91D6B4E2BF CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2226;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
```

```
Db    249 NEFRV 253


RESULT  10
POLG_HPAV4
ID   POLG_HPAV4     STANDARD;      PRT;  2226 AA.
AC   P26581;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain 43c).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91162758.
RA   Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA   Cromeans T., Jansen R.W.;
RT   "Antigenic and genetic variation in cytopathic hepatitis A virus
RT   variants arising during persistent infection: evidence for genetic
RT   recombination.";
RL   J. Virol. 65:2056-2065(1991).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59809; AAA45469.1; -.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    794       COAT PROTEIN VP1 (P1D).
FT   CHAIN       795    900       CORE PROTEIN P2A.
FT   CHAIN       901   1087       CORE PROTEIN P2B.
FT   CHAIN      1088   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1495       PROBABLE PROTEIN P3A.
FT   CHAIN      1496   1518       PROBABLE PROTEIN P3B.
FT   CHAIN      1519   1737       PROBABLE PROTEIN P3C.
FT   CHAIN      1738   2226       RNA-DIRECTED POLYMERASE P3D.
SQ   SEQUENCE   2226 AA;  251107 MW;  403B4CA80B09BF75 CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2226;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  11
POLG_HPAV8
ID   POLG_HPAV8     STANDARD;      PRT;  2226 AA.
AC   P26582;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain 18f).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91162758.
RA   Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA   Cromeans T., Jansen R.W.;
RT   "Antigenic and genetic variation in cytopathic hepatitis A virus
RT   variants arising during persistent infection: evidence for genetic
RT   recombination.";
RL   J. Virol. 65:2056-2065(1991).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59808; AAA45467.1; -.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    794       COAT PROTEIN VP1 (P1D).
FT   CHAIN       795    900       CORE PROTEIN P2A.
FT   CHAIN       901   1087       CORE PROTEIN P2B.
FT   CHAIN      1088   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1495       PROBABLE PROTEIN P3A.
FT   CHAIN      1496   1518       PROBABLE PROTEIN P3B.
FT   CHAIN      1519   1737       PROBABLE PROTEIN P3C.
FT   CHAIN      1738   2226       RNA-DIRECTED POLYMERASE P3D.
SQ   SEQUENCE   2226 AA;  251292 MW;  24964A63396C8D6B CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2226;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  12
POLG_HPAVH
ID   POLG_HPAVH     STANDARD;      PRT;  2227 AA.
AC   P08617; P06443; Q81082;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain HM-175).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=WILD TYPE;
RX   MEDLINE; 87061253.
RA   Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,
RA   Baroudy B.M.;
RT   "Complete nucleotide sequence of wild-type hepatitis A virus:
RT   comparison with different strains of hepatitis A virus and other
RT   picornaviruses.";
RL   J. Virol. 61:50-59(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATTENUATED;
RX   MEDLINE; 87175701.
RA   Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,
RA   Purcell R.H.;
RT   "Complete nucleotide sequence of an attenuated hepatitis A virus:
RT   comparison with wild-type virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).
RN   [3]
RP   SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.
RX   MEDLINE; 85166289.
RA   Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,
RA   Purcell R.H., Feinstone S.M.;
RT   "Sequence analysis of hepatitis A virus cDNA coding for capsid
RT   proteins and RNA polymerase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
CC       FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   -!- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT
CC       SHOWN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14114; AAA45475.1; -.
DR   EMBL; M14707; AAA45465.1; -.
DR   EMBL; M14707; AAA45466.1; ALT_INIT.
DR   EMBL; M16632; AAA45471.1; -.
DR   PIR; A25981; GNNYHM.
DR   PIR; A25914; GNNYMK.
DR   PIR; A03905; A03905.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    836       COAT PROTEIN VP1 (P1D).
FT   CHAIN       837    980       CORE PROTEIN P2A.
FT   CHAIN       981   1087       CORE PROTEIN P2B.
FT   CHAIN      1088   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1496       PROBABLE PROTEIN P3A.
FT   CHAIN      1497   1519       PROBABLE PROTEIN P3B.
FT   CHAIN      1520   1738       PROBABLE PROTEIN P3C.
FT   CHAIN      1739   2227       RNA-DIRECTED POLYMERASE 3D.
FT   VARIANT      77     77       K -> R (IN ATTENUATED STRAIN).
FT   VARIANT     764    764       E -> V (IN ATTENUATED STRAIN).
FT   VARIANT     821    821       N -> S (IN ATTENUATED STRAIN).
FT   VARIANT    1052   1052       A -> V (IN ATTENUATED STRAIN).
FT   VARIANT    1062   1062       G -> A (IN ATTENUATED STRAIN).
FT   VARIANT    1118   1118       K -> M (IN ATTENUATED STRAIN).
FT   VARIANT    1151   1151       E -> K (IN ATTENUATED STRAIN).
FT   VARIANT    1163   1163       F -> S (IN ATTENUATED STRAIN).
FT   VARIANT    1277   1277       V -> I (IN ATTENUATED STRAIN).
FT   VARIANT    1500   1500       H -> Y (IN ATTENUATED STRAIN).
FT   VARIANT    1805   1805       D -> N (IN ATTENUATED STRAIN).
FT   VARIANT    1930   1930       S -> T (IN ATTENUATED STRAIN).
SQ   SEQUENCE   2227 AA;  251506 MW;  01E225E7AEB740A6 CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  13
POLG_HPAVL
ID   POLG_HPAVL     STANDARD;      PRT;  2227 AA.
AC   P06441;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain LA).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85190549.
RA   Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA   Merryweather J., van Nest G., Dina D.;
RT   "Primary structure and gene organization of human hepatitis A virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is  opyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02990; AAA45472.1; -.
DR   PIR; A03903; GNNYHR.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    836       COAT PROTEIN VP1 (P1D).
FT   CHAIN       837    980       CORE PROTEIN P2A.
FT   CHAIN       981   1076       CORE PROTEIN P2B.
FT   CHAIN      1077   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1484       PROBABLE PROTEIN P3A.
FT   CHAIN      1485   1507       PROBABLE PROTEIN P3B.
FT   CHAIN      1508   1678       PROBABLE PROTEIN P3C.
FT   CHAIN      1679   2227       RNA-DIRECTED POLYMERASE P3D.
SQ   SEQUENCE   2227 AA;  251898 MW;  99A7354B4CD2799C CRC64;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  14
POLG_HPAVM
ID   POLG_HPAVM     STANDARD;      PRT;  2227 AA.
AC   P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
AC   Q81090; Q81091; Q81092; Q81093;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Hepatitis A virus (strain MBB).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88045071.
RA   Paul A.V., Tada H., der Helm K., Wissel T., Kiehn R., Wimmer E.,
RA   Deinhardt F.;
RT   "The entire nucleotide sequence of the genome of human hepatitis A
RT   virus (isolate MBB).";
RL   Virus Res. 8:153-171(1987).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT
CC       WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20273; AAA45474.1; -.
DR   PIR; JS0303; GNNYHB.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     23       COAT PROTEIN VP4 (P1A).
FT   CHAIN        24    245       COAT PROTEIN VP2 (P1B).
FT   CHAIN       246    491       COAT PROTEIN VP3 (P1C).
FT   CHAIN       492    836       COAT PROTEIN VP1 (P1D).
FT   CHAIN       837    980       CORE PROTEIN P2A.
FT   CHAIN       981   1087       CORE PROTEIN P2B.
FT   CHAIN      1088   1422       CORE PROTEIN P2C.
FT   CHAIN      1423   1496       PROBABLE PROTEIN P3A.
FT   CHAIN      1497   1519       PROBABLE PROTEIN P3B.
FT   CHAIN      1520   1738       PROBABLE PROTEIN P3C.
FT   CHAIN      1739   2227       RNA-DIRECTED POLYMERASE 3D.
SQ   SEQUENCE   2227 AA;  251425 MW;  EC983ED2A7C8C349 CRC64;


  Query Match           100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  15
POLG_HPAVS
ID   POLG_HPAVS     STANDARD;      PRT;  2230 AA.
AC   P14553;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE   P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE   P3D (EC 2.7.7.48)].
OS   Simian hepatitis A virus (strain AGM-27).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91311420.
RA   Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA   Purcell R.H.;
RT   "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT   structure and growth in cell culture with other HAV strains.";
RL   J. Gen. Virol. 72:1677-1683(1991).
RN   [2]
RP   SEQUENCE OF 1750-2164 FROM N.A.
RX   MEDLINE; 89232168.
RA   Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
RA   Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RT   "Variations in genome fragments coding for RNA polymerase in human
RT   and simian hepatitis A viruses.";
RL   FEBS Lett. 247:425-428(1989).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00924; BAA00766.1; -.
DR   EMBL; X15461; CAA33490.1; -.
DR   PIR; A30470; GNNYSA.
DR   PIR; S04885; S04885.
DR   PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR   PFAM; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         1     27       COAT PROTEIN VP4 (P1A).
FT   CHAIN        28    249       COAT PROTEIN VP2 (P1B).
FT   CHAIN       250    495       COAT PROTEIN VP3 (P1C).
FT   CHAIN       496    795       COAT PROTEIN VP1 (P1D).
FT   CHAIN       796    984       CORE PROTEIN P2A.
FT   CHAIN       985   1091       CORE PROTEIN P2B.
FT   CHAIN      1092   1426       CORE PROTEIN P2C.
FT   CHAIN      1427   1498       PROBABLE PROTEIN 3A.
FT   CHAIN      1499   1521       PROBABLE PROTEIN 3B.
FT   CHAIN      1522   1741       PROBABLE PROTEIN 3C.
FT   CHAIN      1742   2230       RNA-DIRECTED POLYMERASE 3D.
SQ   SEQUENCE   2230 AA;  251296 MW;  87B3230E324E1F19 CRC64;


  Query Match           100.0%;  Score 26;  DB 1;  Length 2230;
  Best Local Similarity  100.0%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     253 NEFRV 257
```

Search completed: June 22, 2000, 19:33:57
Job time: 1100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:30:01 ; Search time 54.21 Seconds
(without alignments)
6.395 Million cell updates/sec

Title: US-09-362-485-29
Perfect score: 26
Sequence: 1 NEFRV 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	74	2	O32280	O32280 bacillus su
2	26	100.0	113	12	Q86532	Q86532 hepatitis a
3	26	100.0	156	5	Q23362	Q23362 caenorhabdi
4	26	100.0	158	11	Q64367	Q64367 mus musculu
5	26	100.0	300	2	O53943	O53943 mycobacteri
6	26	100.0	300	2	Q9Z5I1	Q9z5i1 mycobacteri
7	26	100.0	343	5	Q9XWH0	Q9xwh0 caenorhabdi
8	26	100.0	369	5	O96121	O96121 plasmodium
9	26	100.0	376	10	Q9XG54	Q9xg54 lycopersico
10	26	100.0	515	5	Q27699	Q27699 musca domes
11	26	100.0	516	5	Q27698	Q27698 musca domes
12	26	100.0	516	5	Q27700	Q27700 musca domes
13	26	100.0	516	5	Q27697	Q27697 musca domes
14	26	100.0	516	5	Q9XYB5	Q9xyb5 musca domes
15	26	100.0	611	10	Q9ZU34	Q9zu34 arabidopsis
16	26	100.0	720	5	Q18437	Q18437 caenorhabdi
17	26	100.0	791	12	Q87095	Q87095 simian hepa
18	26	100.0	1067	12	P89904	P89904 chimpanzee
19	26	100.0	1124	12	Q84780	Q84780 hepatitis a
20	26	100.0	1160	5	Q27829	Q27829 paramecium
21	26	100.0	1161	12	Q05794	Q05794 hepatitis a
22	26	100.0	2216	12	Q9WMA2	Q9wma2 hepatitis a
23	26	100.0	2218	12	Q67824	Q67824 hepatitis a
24	26	100.0	2218	12	Q67817	Q67817 hepatitis a
25	26	100.0	2227	12	Q67825	Q67825 hepatitis a
26	26	100.0	2227	12	Q67826	Q67826 hepatitis a
27	26	100.0	2227	12	Q9WMA4	Q9wma4 hepatitis a
28	26	100.0	2227	12	Q9WMA3	Q9wma3 hepatitis a
29	26	100.0	2227	12	Q9WMA1	Q9wma1 hepatitis a
30	26	100.0	2227	12	Q9WMA0	Q9wma0 hepatitis a
31	26	100.0	2227	12	Q9WM99	Q9wm99 hepatitis a
32	25	96.2	188	8	O21034	O21034 dictyosteli
33	25	96.2	306	12	Q9YWQ0	Q9ywq0 coltivirus
34	25	96.2	306	12	Q9YWN3	Q9ywn3 coltivirus
35	25	96.2	343	5	Q19781	Q19781 caenorhabdi
36	25	96.2	592	9	Q9ZXI4	Q9zxi4 pseudomonas
37	25	96.2	637	5	O61730	O61730 drosophila
38	25	96.2	1483	5	P90865	P90865 caenorhabdi
39	23	88.5	35	2	O87902	O87902 coxiella bu
40	23	88.5	49	12	O56636	O56636 hepatitis c
41	23	88.5	81	6	O97893	O97893 sus scrofa
42	23	88.5	89	1	O06113	O06113 methanococc
43	23	88.5	103	2	Q9ZFF5	Q9zff5 legionella
44	23	88.5	103	2	Q9ZFF4	Q9zff4 legionella
45	23	88.5	104	2	Q9ZFF3	Q9zff3 legionella

ALIGNMENTS

RESULT 1
O32280
ID O32280 PRELIMINARY; PRT; 74 AA.
AC O32280;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE YWZC PROTEIN.
GN YWZC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE B., O'REILLY M., OGASAWARA A., OUDEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,

```
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99123; CAB15788.1; -.
SQ   SEQUENCE   74 AA;  8327 MW;  AFE27F7B CRC32;
```

Query Match 100.0%; Score 26; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |||||
Db       3 NEFRV 7
```

```
RESULT   2
Q86532
ID   Q86532      PRELIMINARY;      PRT;   113 AA.
AC   Q86532;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   VP3 PROTEIN (FRAGMENT).
GN   VP3.
OS   Hepatitis A virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92348853.
RA   FINESCHI N., CAVALIERI F., GARELICK H., PRUGNOLA A., PELLEGRINI V.,
RA   ZUCKERMAN A.J.;
RT   "Characterization of a hepatitis A virus strain suitable for vaccine
RT   production.";
RL   J. Hepatol. 13:51-51(1991).
DR   EMBL; S44086; AAB22737.1; -.
FT   NON_TER     113    113
SQ   SEQUENCE   113 AA;  12812 MW;  AC7CDABC CRC32;
```

Query Match 100.0%; Score 26; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |||||
Db       4 NEFRV 8
```

```
RESULT   3
Q23362
ID   Q23362      PRELIMINARY;      PRT;   156 AA.
AC   Q23362;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CODED FOR BY C. ELEGANS CDNA CEESY85F.
GN   ZC513.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WU X., LE T.T.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53155; AAC48269.1; -.
SQ   SEQUENCE   156 AA;  17779 MW;  38974034 CRC32;
```

Query Match 100.0%; Score 26; DB 5; Length 156;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |||||
Db     129 NEFRV 133
```

```
RESULT   4
Q64367
ID   Q64367      PRELIMINARY;      PRT;   158 AA.
AC   Q64367;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SEMINAL VESICLE ANTIGEN PRECURSOR (SEMINAL VESICLE AUTOANTIGEN).
GN   SVA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57 BLACK/6XCBA; TISSUE=SPLEEN;
RA   YU L.C., HSIAO Y.L., YANG Y.H., CHEN Y.H.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94107218.
RA   YU L.C., CHEN J.L., TSAI W.B., CHEN Y.H.;
RT   "Primary structure and characterization of an androgen-stimulated
RT   autoantigen purified from mouse seminal-vesicle secretion.";
RL   Biochem. J. 296:571-576(1993).
DR   EMBL; L44117; AAA73510.1; -.
DR   EMBL; M94179; AAA16593.1; -.
DR   MGD; MGI:102785; Sva.
KW   Signal; Seminal vesicle.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    158       SEMINAL VESICLE AUTOANTIGEN.
SQ   SEQUENCE   158 AA;  18266 MW;  CF5918FE CRC32;
```

Query Match 100.0%; Score 26; DB 11; Length 158;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |||||
Db      60 NEFRV 64
```

```
RESULT   5
```

```
O53943
ID   O53943      PRELIMINARY;      PRT;   300 AA.
AC   O53943;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 32.4 KD PROTEIN.
GN   MTV049.16.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   SKELTON J., CHURCHER C.M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; AL022021; CAA17715.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   300 AA;  32399 MW;  B4F83856 CRC32;
```

```
  Query Match             100.0%;  Score 26;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     114 NEFRV 118
```

```
RESULT   6
Q9Z5I1
ID   Q9Z5I1      PRELIMINARY;      PRT;   300 AA.
AC   Q9Z5I1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 32.6 KD PROTEIN.
GN   MLCB596.31.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   OLIVER K., HARRIS D.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   JAMES K.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93188700.
RA   EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Mol. Microbiol. 7:197-206(1993).
DR   EMBL; AL035472; CAB36587.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   300 AA;  32550 MW;  1E26BF74 CRC32;
```

```
  Query Match             100.0%;  Score 26;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     114 NEFRV 118
```

```
RESULT   7
Q9XWH0
ID   Q9XWH0      PRELIMINARY;      PRT;   343 AA.
AC   Q9XWH0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   Y54G9A.6 PROTEIN.
GN   Y54G9A.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   SMYE R.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; AL032648; CAA21698.1; -.
SQ   SEQUENCE   343 AA;  38127 MW;  31657AB0 CRC32;
```

```
  Query Match             100.0%;  Score 26;  DB 5;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db      14 NEFRV 18
```

```
RESULT   8
O96121
ID   O96121      PRELIMINARY;      PRT;   369 AA.
AC   O96121;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 45.0 KD PROTEIN.
GN   PFB0080C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99021743.
RA   GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA   KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA   SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA   SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA   FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
DR   EMBL; AE001369; AAC71806.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   369 AA;  44979 MW;  5B18C387 CRC32;

  Query Match             100.0%;  Score 26;  DB 5;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     154 NEFRV 158


RESULT   9
Q9XG54
ID   Q9XG54      PRELIMINARY;      PRT;   376 AA.
AC   Q9XG54;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   12-OXOPHYTODIENOATE REDUCTASE (EC 1.3.1.42).
GN   OPR.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC   Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CASTLEMART II; TISSUE=SHOOT;
RA   STRASSNER J., FURHOLZ A., MACHEROUX P., AMRHEIN N., SCHALLER A.;
RT   "A homolog of Old Yellow Enzyme in tomato: Spectral properties and
RT   substrate specificity of the recombinant protein.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ242551; CAB43506.1; -.
KW   Oxidoreductase.
SQ   SEQUENCE   376 AA;  42415 MW;  D5A456EC CRC32;

  Query Match             100.0%;  Score 26;  DB 10;  Length 376;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     167 NEFRV 171


RESULT  10
Q27699
ID   Q27699      PRELIMINARY;      PRT;   515 AA.
AC   Q27699;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PYRETHROID RESISTANCE CYTOCHROME P450 (EC 1.14.14.1).
GN   CYP6D1.
OS   Musca domestica (House fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Muscoidea; Muscidae; Musca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AABYS;
RX   MEDLINE; 96111292.
RA   TOMITA T., LIU N., SMITH F.F., SRIDHAR P., SCOTT J.G.;
RT   "Molecular mechanisms involved in increased expression of a cytochrome
RT   P450 responsible for pyrethroid resistance in the housefly, Musca
RT   domestica.";
RL   Insect Mol. Biol. 4:135-140(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AABYS;
RA   SCOTT J.G., LIU N., WEN Z., SMITH F.F., HORAK C.E.;
RT   "House fly cytochrome P450 CYP6D1: 5' flanking sequences and
RT   comparison of two alleles.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U22366; AAC46930.1; -.
DR   EMBL; AF064795; AAC99341.1; -.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PFAM; PF00067; p450; 1.
DR   PRINTS; PR00464; EP450II.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   BINDING     460    460       HEME (BY SIMILARITY).
SQ   SEQUENCE   515 AA;  58982 MW;  E2E3A87F CRC32;

  Query Match             100.0%;  Score 26;  DB 5;  Length 515;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     198 NEFRV 202


RESULT  11
Q27698
ID   Q27698      PRELIMINARY       PRT;   516 AA.
AC   Q27698;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME P450 (EC 1.14.14.1).
GN   CYP6D1.
OS   Musca domestica (House fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Muscoidea; Muscidae; Musca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CS;
RX   MEDLINE; 96111292.
RA   TOMITA T., LIU N., SMITH F.F., SRIDHAR P., SCOTT J.G.;
RT   "Molecular mechanisms involved in increased expression of a cytochrome
RT   P450 responsible for pyrethroid resistance in the housefly, Musca
RT   domestica.";
RL   Insect Mol. Biol. 4:135-140(1995).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U22362; AAC46932.1; -.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PFAM; PF00067; p450; 1.
DR   PRINTS; PR00464; EP450II.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   BINDING     461    461       HEME (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  59172 MW;  7339B054 CRC32;

  Query Match             100.0%;  Score 26;  DB 5;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
```

```
          |||||
Db    199 NEFRV 203


RESULT  12
Q27700
ID   Q27700      PRELIMINARY;      PRT;   516 AA.
AC   Q27700;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PYRETHROID RESISTANCE CYTOCHROME P450 (EC 1.14.14.1).
GN   CYP6D1.
OS   Musca domestica (House fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Muscoidea; Muscidae; Musca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ISK;
RX   MEDLINE; 96111292.
RA   TOMITA T., LIU N., SMITH F.F., SRIDHAR P., SCOTT J.G.;
RT   "Molecular mechanisms involved in increased expression of a cytochrome
RT   P450 responsible for pyrethroid resistance in the housefly, Musca
RT   domestica.";
RL   Insect Mol. Biol. 4:135-140(1995).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U22367; AAC46931.1; -.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PFAM; PF00067; p450; 1.
DR   PRINTS; PR00464; EP450II.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   BINDING     461    461       HEME (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  59154 MW;  85FE0324 CRC32;


  Query Match             100.0%;  Score 26;  DB 5;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     199 NEFRV 203


RESULT  13
Q27697
ID   Q27697      PRELIMINARY;      PRT;   516 AA.
AC   Q27697;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME P450 (EC 1.14.14.1).
GN   CYP6D1.
OS   Musca domestica (House fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Muscoidea; Muscidae; Musca.
RN   [1]
RP   SEQUENCE OF 1-514 FROM N.A.
RC   STRAIN=LEARN-PYR;
RX   MEDLINE; 95227198.
RA   TOMITA T., SCOTT J.G.;
RT   "cDNA and deduced protein sequence of CYP6D1: the putative gene for a
RT   cytochrome P450 responsible for pyrethroid resistance in house fly.";
RL   Insect Biochem. Mol. Biol. 25:275-283(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEARN-PYR;
RX   MEDLINE; 95180368.
RA   LIU N., TOMITA T., SCOTT J.G.;
RT   "Allele-specific PCR reveals that CYP6D1 is on chromosome 1 in the
RT   house fly, Musca domestica.";
RL   Experientia 51:164-167(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEARN-PYR;
RA   SCOTT J.G.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LPR;
RA   SCOTT J.G., LIU N., WEN Z., SMITH F.F., HORAK C.E.;
RT   "House fly cytochrome P450 CYP6D1: 5' flanking sequences and
RT   comparison of two alleles.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U15168; AAA81513.1; -.
DR   EMBL; AF064794, AAC99340.1; -.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PFAM; PF00067; p450; 1.
DR   PRINTS; PR00464; EP450II.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   BINDING     461    461       HEME (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  59184 MW;  6362082A CRC32;


  Query Match             100.0%;  Score 26;  DB 5;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     199 NEFRV 203


RESULT  14
Q9XYB5
ID   Q9XYB5      PRELIMINARY;      PRT;   516 AA.
AC   Q9XYB5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME P450.
GN   CYP6D1.
OS   Musca domestica (House fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Muscoidea; Muscidae; Musca.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OCR;
RA   SCOTT J.G., LIU N., WEN Z., SMITH F.F., KASAI S., HORAK C.E.;
RT   "House fly cytochrome P450 CYP6D1: 5' flanking sequences and
RT   comparison of alleles.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF081288; AAD32478.1; -.
SQ   SEQUENCE   516 AA;  59115 MW;  6A5D9BF1 CRC32;


  Query Match             100.0%;  Score 26;  DB 5;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     199 NEFRV 203


RESULT  15
Q9ZU34
ID   Q9ZU34      PRELIMINARY;      PRT;   611 AA.
AC   Q9ZU34;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
```

DE PUTATIVE STRESS PROTEIN.
GN F10A8.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA FRASER C.M., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC F10A8 genomic sequence.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC006200; AAD14533.1; -.
SQ SEQUENCE 611 AA; 66387 MW; 3E6A3432 CRC32;

Query Match 100.0%; Score 26; DB 10; Length 611;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 94 NEFRV 98

Search completed: June 22, 2000, 19:30:03
Job time: 5145 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 09:59:45 ; Search time 8627.09 Seconds
(without alignments)
-142.078 Million cell updates/sec

Title: US-09-362-485-1
Perfect score: 1260
Sequence: 1 GAATTCCCATCAGCAATCTT.....AAGCGATGATGTCGGCCGCG 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1260	100.0	1260	5	A87604	A87604 Sequence 1
2		1260	100.0	1260	5	A89744	A89744 Sequence 1
3		1260	100.0	2584	1	MTALADH	X63069 M.tuberculo
4		1245	98.8	1245	5	A87605	A87605 Sequence 2
5		1245	98.8	1245	5	A89745	A89745 Sequence 2
6		1233.2	97.9	56414	1	MTV002	AL008967 Mycobacte
7		1220	96.8	1236	5	A87613	A87613 Sequence 10
8		1220	96.8	1236	5	A89753	A89753 Sequence 10
9		1219.4	96.8	1237	5	A87607	A87607 Sequence 4
10		1219.4	96.8	1237	5	A89747	A89747 Sequence 4
11		1216	96.5	123[illegible]	5	A87606	A87606 Sequence 3
12		1216	96.5	123[illegible]	5	A89746	A89746 Sequence 3
13		1210.4	96.1	1228	5	A87608	A87608 Sequence 5
14		1210.4	96.1	1228	5	A89748	A89748 Sequence 5
15		1207	95.8	1235	5	A87609	A87609 Sequence 6
16		1207	95.8	1235	5	A89749	A89749 Sequence 6
17		1206.4	95.7	1235	5	A87611	A87611 Sequence 8
18		1206.4	95.7	1235	5	A89751	A89751 Sequence 8
19		1199.4	95.2	1229	5	A87610	A87610 Sequence 7
20		1199.4	95.2	1229	5	A89750	A89750 Sequence 7
21		1192	94.6	1209	5	A87612	A87612 Sequence 9
22		1185	94.0	1208	5	A89752	A89752 Sequence 9
23		1178	93.5	1194	2	MTU92472	U92472 Mycobacteri
24		720	57.1	720	5	A89759	A89759 Sequence 16
25	c	522.4	41.5	40745	1	SCI51	AL109848 Streptomy
26		467.6	37.1	682	5	A89758	A89758 Sequence 15
27		467.6	37.1	682	5	A89762	A89762 Sequence 19
28		361	28.7	8046	2	AF049107	AF049107 Myxococcu
29		337.2	26.8	11514	2	AE002028	AE002028 Deinococc
30	c	315.4	25.0	2780	1	RLE238118	AJ238118 Rhizobium
31		314.8	25.0	1585	2	AF070716	AF070716 Vibrio pr
32		296.8	23.6	1736	1	BACALDHC	M33299 B.stearothe
33		277	22.0	1477	2	AF070715	AF070715 Shewanell
34		272.4	21.6	2113	1	D37807	D37807 Phormidium
35		261.2	20.7	1125	5	E01847	E01847 DNA encodin
36	c	246.8	19.6	137740	1	D90900	D90900 Synechocyst
37		227.2	18.0	1375	1	AB013821	AB013821 Enterobac
38		227	18.0	1776	1	BACALDH	M33298 B.sphaericu
39		219.6	17.4	1709	1	BACALD	L20916 Bacillus su
40		219.6	17.4	12316	1	BSZ82015	Z82015 B.subtilis
41		219.6	17.4	217420	1	BSUB0017	Z99120 Bacillus su
42		202.2	16.0	1714	2	AF070714	AF070714 Carnobact
43		108.8	8.6	10652	2	AE001565	AE001565 Helicobac
44	c	100.8	8.0	10512	2	AE000639	AE000639 Helicobac
45		66.4	5.3	3160	1	RRU05294	U05294 Rhodospiril

ALIGNMENTS

```
RESULT   1
A87604
LOCUS       A87604       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9836089.
ACCESSION   A87604
VERSION     A87604.1  GI:6736244
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE.
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN

  Query Match           100.0%;  Score 1260;  DB 5;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 3.9e-174;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60

Qy      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120

Qy     121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180

Qy     181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240

Qy     241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300

Qy     301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360

Qy     361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420

Qy     421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480

Qy     481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540

Qy     541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600

Qy     601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660

Qy     661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720

Qy     721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780

Qy     781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840

Qy     841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900

Qy     901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960

Qy     961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACGTACGCGCTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020

Qy    1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080

Qy    1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140

Qy    1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200

Qy    1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TCGGCCGCTCGTTACGCC AGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260


RESULT   2
A89744
LOCUS       A89744       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9832862.
ACCESSION   A89744
VERSION     A89744.1  GI:6738278
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      243 a    403 c    389 g    225 t
ORIGIN

  Query Match           100.0%;  Score 1260;  DB 5;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 3.9e-174;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60

Qy      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
Qy    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
Db    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
Qy    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
Db    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
Qy    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
Db    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
Qy    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
Db    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
Qy    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
Db    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
Qy    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
Db    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
Qy    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
Db    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
Qy    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCCAACCGGCCGACGTCGTGGTG 600
Db    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
Qy    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
Db    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
Qy    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
Db    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
Qy    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
Db    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
Qy    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
Db    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
Qy    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
Db    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
Qy    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
Db    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
Qy    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
Db    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
Qy   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
Db   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
Qy   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
Db   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
Qy   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
Db   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
Qy   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
Db   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
```

```
RESULT   3
MTALADH
LOCUS       MTALADH      2584 bp    DNA             BCT       01-DEC-1993
DEFINITION  M.tuberculosis gene for L-alanine dehydrogenase.
ACCESSION   X63069 S36765
VERSION     X63069.1  GI:44565
KEYWORDS    alanine dehydrogenase; extracellular; intracellular.
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 2584)
  AUTHORS   Andersen,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1991) A.B. Andersen, Statens Seruminstitut,
            Artillerivej 5, DK-2300 Copenhagen S, DENMARK
REFERENCE   2  (bases 1 to 2584)
  AUTHORS   Andersen,A.B., Andersen,P. and Ljungqvist,L.
  TITLE     Structure and function of a 40,000-molecular-weight protein antigen
            of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 60 (6), 2317-2323 (1992)
  MEDLINE   92267644
FEATURES             Location/Qualifiers
     source          1. .2584
                     /organism="Mycobacterium tuberculosis"
                     /strain="Erdman"
                     /isolate="TMC 107, lot 9A-2"
                     /db_xref="taxon:1773"
                     /clone_lib="lambda gt11"
                     /clone="lambda AA67"
     CDS             76. .1197
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="alanine  dehydrogenase"
                     /protein_id="CAA44791.1"
                     /db_xref="GI:44566"
                     /db_xref="SWISS-PROT:P30234"
                     /translation="MRVGIPTETKNNEFQFRVAITPAGVAELTRRGHEVLIQAGAGEG
                     SAITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAAS
                     RACTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGV
                     LMGGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIH
                     TRYSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGC
                     FEGSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAA
                     CRSNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      463 a    871 c    819 g    431 t
ORIGIN
```

```
  Query Match           100.0%;  Score 1260;  DB 1;  Length 2584;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-174;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
Db      1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
Qy     61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
Db     61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
Qy    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
Db    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
Qy    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
```

```
Db    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
Qy    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
Db    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
Qy    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
Db    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
Qy    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
Db    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
Qy    421 TCAATTGCCTACGAGACGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
Db    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
Qy    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTT..CCACCTGATGCGAACCCAA 540
Db    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
Qy    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
Db    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
Qy    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
Db    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
Qy    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
Db    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
Qy    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
Db    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
Qy    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
Db    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
Qy    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
Db    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
Qy    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
Db    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
Qy    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
Db    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
Qy   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
Db   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
Qy   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
Db   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
Qy   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
Db   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
Qy   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
Db   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260

RESULT 4
```

```
A87605
LOCUS       A87605       1245 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9836089.
ACCESSION   A87605
VERSION     A87605.1  GI:6736245
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN
```

Query Match 98.8%; Score 1245; DB 5; Length 1245;
Best Local Similarity 100.0%; Pred. No. 5.8e-172;
Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
Db      1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60
Qy     76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
Db     61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 120
Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
Db    121 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 180
Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
Db    181 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 240
Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
Db    241 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 300
Qy    316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
Db    301 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 360
Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
Db    361 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 420
Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
Db    421 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 480
Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
Db    481 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 540
Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
Db    541 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 600
Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
Db    601 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 660
Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
Db    661 ATCAAC TCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 720
```

```
Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 780

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 840

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 900

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 960

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1020

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1080

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1140

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1200

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1245
```

```
RESULT   5
A89745
LOCUS       A89745       1245 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9832862.
ACCESSION   A89745
VERSION     A89745.1  GI:6738279
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN
```

```
  Query Match             98.8%;  Score 1245;  DB 5;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-172;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 120
```

```
Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 180

Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 240

Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 300

Qy    316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 360

Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 420

Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 480

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 540

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 600

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 660

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 720

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 780

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 840

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 900

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 960

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1020

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1080

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1140

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1200

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
```

```
                     ||||||||||||||||||||||||||||||||||||||||
Db    1201 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1245


RESULT   6
MTV002
LOCUS       MTV002      56414 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 122/162.
ACCESSION   AL008967 AL123456
VERSION     AL008967.1  GI:3261491
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 56414)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 56414)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2624256.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .56414
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     misc_feature    <1. .56414
                     /note="fragment designated v002. Does not represent a
                     physical clone"
     source          <1. .233
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y154"
     gene            complement(3. .527)
                     /gene="recX"
     CDS             complement(3. .527)
                     /gene="recX"
                     /note="Rv2736c, (MTV002.01c), len: 174 aa. recX, similar
                     to eg. RECX_PSEAE P37860 regulatory protein recx from
                     Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161
                     z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
                     overlap. Overlaps and extends CDS from ovelapping cosmid
                     MTCY154.16c"
                     /codon_start=1
                     /transl_table=11
                     /product="recX"
                     /protein_id="CAA15532.1"
                     /db_xref="GI:2624257"
                     /db_xref="SPTREMBL:O33280"
                     /translation="MTVSCPPPSTSEREEQARALCLRLLTARSRTRAELAGQLAKRGY
                     PEDIGNRVLDRLAAVGLVDDTDFAEQWVQSRRANAAKSKRALAAELHAKGVDDDVITT
                     VLGGIDAGAERGRAEKLVRARLRREVLIDDGTDEARVSRRLVAMLARRGYGQTLACEV
                     VIAELAAERERRRV"
     gene            complement(493. .2865)
                     /gene="recA"
     CDS             complement(493. .2865)
                     /gene="recA"
                     /note="Rv2737c, (MTV002.02c), len: 790 aa. recA,
                     identicalto RECA_MYCTU P26345 reca protein (790 aa).
                     Contains self-splicing protein element (intein) from 2294
                     to 969 (c), similar to intein II from TR:E332317
                     (EMBL:Y13030) DNA-directedDNA polymerase (EC 2.7.7.7) from
                     Thermococcus sp. (1829 aa), fasta scores; opt: 81 z-score:
                     235.2 E(): 6e-06, 24.6%identity in 183 aa overlap.
                     Contains PS00017 ATP/GTP-bindingsite motif A (P-loop),
                     PS00321 recA signature, and PS00881Protein splicing
                     signature. See Davis et al, (1992) Cell71(2):201-210"
                     /codon_start=1
                     /transl_table=11
                     /product="recA"
                     /protein_id="CAA15533.1"
                     /db_xref="GI:2624258"
                     /db_xref="SWISS-PROT:P26345"
                     /translation="MTQTPDREKALELAVAQIEKSYGKGSVMRLGDEARQPISVIPTG
                     SIALDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGVAAFIDAEHALDP
                     DYAKKLGVDTDSLLVSQPDTGEQALEIADMLIRSGALDIVVIDSVAALVPRAELEGEM
                     GDSHVGLQARLMSQALRKMTGALNNSGTTAIFINQLRDKIGVMFGSPETTTGGKALKF
                     YASVRMDVRRVETLKDGTNAVGNRTRVKVVKNKCLAEGTRIFDPVTGTTHRIEDVVDG
                     RKPIHVVAAAKDGTLHARPVVSWFDQGTRDVIGLRIAGGAIVWATPDHKVLTEYGWRA
                     AGELRKGDRVAQPRRFDGFGDSAPIPADHARLLGYLIGDGRDGWVGGKTPINFINVQR
                     ALIDDVTRIAATLGCAAHPQGRISLAIAHRPGERNGVADLCQQAGIYGKLAWEKTIPN
                     WFFEPDIAADIVGNLLFGLFESDGWVSREQTGALRVGYTTTSEQLAHQIHWLLLRFGV
                     GSTVRDYDPTQKRPSIVNGRRIQSKRQVFEVRISGMDNVTAFAESVPMWGPRGAALIQ
                     AIPEATQGRRRGSQATYLAAEMTDAVLNYLDERGVTAQEAAAMIGVASGDPRGGMKQV
                     LGASRLRRDRVQALADALDDKFLHDMLAEELRYSVIREVLPTRRARTFDLEVEELHTL
                     VAEGVVVHNCSPPFKQAEFDILYGKGISREGSLIDMGVDQGLIRKSGAWFTYEGEQLG
                     QGKENARNFLVENADVADEIEKKIKEKLGIGAVVTDDPSNDGVLPAPVDF"
     misc_feature    complement(790. .2112)
                     /gene="recA"
                     /note="recA intein"
     misc_feature    complement(790. .807)
                     /gene="recA"
                     /note="PS00881 Protein splicing signature"
     misc_feature    complement(2197. .2223)
                     /gene="recA"
                     /note="PS00321 recA signature"
     misc_feature    complement(2644. .2667)
                     /gene="recA"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     RBS             complement(2873. .2877)
                     /note="possible RBS upstream of Rv2737c"
     gene            complement(3247. .3453)
                     /gene="Rv2738c"
     CDS             complement(3247. .3453)
                     /gene="Rv2738c"
                     /note="Rv2738c, (MTV002.03c), len: 68 aa. Unknown,
                     N-terminus is highly similar to the N-terminus of the
                     upstream ORFMTV002.07c (78.4% identity in 37 aa overlap);
                     also similarto AL020958|SC4H8_5 Streptomyces coelicolor
```

```
                cosmid 4H8 (64aa)  opt:  185 z-score: 283.5 E(): 2.9e-08;
                39.7% identityin 63 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2738c"
                /protein_id="CAA15534.1"
                /db_xref="GI:2624260"
                /db_xref="SPTREMBL:O33281"
                /translation="MLAGVRLTEFHERVALHFGAAYGSSVLLDHVLTGFDGRSAAQAI
                EDGVEPRDVWRALCADFDVPHDRW"
     gene       complement(3464. .4630)
                /gene="Rv2739c"
     CDS        complement(3464. .4630)
                /gene="Rv2739c"
                /note="Rv2739c, (MTV002.04c), len: 388 aa. Probable
                transferase, similar to eg. TR:Q51560 (EMBL:L28170)
                Pseudomonas aeruginosa rhamnosyl transferase (426 aa),
                fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
                identity in 425 aaoverlap. Equivalent to Mycobacterium
                leprae protein MLCB33.02c (392 aa); fasta scores
                gp|Z94723|MLCB33_2 Mycobacteriumleprae cosmid B33 opt:
                2112 z-score: 2364.5 E(): 0; 80.9%identity in 388 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2739c"
                /protein_id="CAA15535.1"
                /db_xref="GI:2624261"
                /db_xref="SPTREMBL:O33282"
                /translation="MRVAVVAGPDPGHSFPAIALCQRFRAAADTPTLFTGVEWLEAAR
                AAGIDAVELDGLAATDRDLDAGARIHRRAAQMAVLNVPRLRALEPELVVSDVITACGG
                MAAELLGIPWVELNPHPLYLPSKGLPPIGSGLAAGTGIRGRLRDATMRALTGRSWRAG
                LRQRAAVRVEIGLPARDPGPLRRLIATLPALEVPRPDWPAEAVVVGPLHFEPTDRVLA
                IPAGTGPVVVVAPSTALTGTAGLTEVALQSLTPGETVPSGSRLVVSRLSGADLTVPPW
                AVAGLGSQAELLTRADLVICGGGHGMVAKTLLAGVPMVVVPGGGDQWEIANRVVRQGS
                AVLIRPLTADALVAAVNEVLSSPRFREAARRAAASVAGAADPVRVCHDALALAG"
     gene       4674. .5123
                /gene="Rv2740"
     CDS        4674. .5123
                /gene="Rv2740"
                /note="Rv2740, (MTV002.05), len: 149 aa. unknown, similar
                to M. leprae hypothetical protein TR:Q49850 (EMBL:Z94723)
                MLCB33.03 (B2235_F3_140) (178 aa), fasta scores; opt: 498
                z-score: 538.9 E(): 7.3e-23, 51.6% identity in 161 aa
                overlap"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2740"
                /protein_id="CAA15536.1"
                /db_xref="GI:2624262"
                /db_xref="SPTREMBL:O33283"
                /translation="MAELTETSPETPETTEAIRAVEAFLNALQNEDFDTVDAALGDDL
                VYENVGFSRIRGGRRTATLLRRMQGRVGFEVKIHRIGADGAAVLTERTDALIIGPLRV
                QFWVCGVFEVDDGRITLWRDYFDVYDMFKGLLRGLVALVVPSLKATL"
     gene       5355. .6932
                /gene="PE_PGRS"
     CDS        5355. .6932
                /gene="PE_PGRS"
                /note="Rv2741, (MTV002.06), len: 525 aa; Member of M.
                tuberculosis PE_ PGRS subfamily, similar to many eg.

  Query Match          97.9%;  Score 1233.2;  DB 1;  Length 56414;
  Best Local Similarity  99.3%;  Pred. No. 1.3e-170;
  Matches 1251;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
          |||||  ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   38186 GAATTATCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCATACTGAAGCGTACAGTA 38245

Qy      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||      ||
Db   38246 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AA 38299
```

```
Qy     121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38300 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 38359

Qy     181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38360 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 38419

Qy     241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38420 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 38479

Qy     301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38480 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 38539

Qy     361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38540 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 38599

Qy     421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38600 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 38659

Qy     481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38660 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 38719

Qy     541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38720 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 38779

Qy     601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38780 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 38839

Qy     661 GTTACGGTTCTAGACATC. .CATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   38840 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 38899

Qy     721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38900 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 38959

Qy     781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38960 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 39019

Qy     841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39020 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 39079

Qy     901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39080 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 39139

Qy     961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39140 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 39199

Qy    1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39200 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 39259

Qy    1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39260 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 39319

Qy    1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39320 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 39379

Qy    1201 TCGGCCCCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
```

```
Db  39380 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 39439


RESULT   7
A87613
LOCUS       A87613       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9836089.
ACCESSION   A87613
VERSION     A87613.1  GI:6736253
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             96.8%;  Score 1220;  DB 5;  Length 1236;
  Best Local Similarity   99.5%;  Pred. No. 2.4e-168;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    6;  Gaps    1;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           ||||||||||||||||||||||||||||||||||||||      ||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy     436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy     496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy     556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCG TGGTGATCGGCGCCGGCACC 594

Qy     616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy     676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy     736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy     796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy     856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy     916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy     976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy    1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy    1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy    1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy    1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1257
           |||||||||||||||||||||||||||||||||||||||||||
Db    1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1236


RESULT   8
A89753
LOCUS       A89753       1236 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 10 from Patent WO9832862.
ACCESSION   A89753
VERSION     A89753.1  GI:6738287
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1236)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1236
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    395 c    385 g    220 t
ORIGIN


  Query Match             96.8%;  Score 1220;  DB 5;  Length 1236;
  Best Local Similarity   99.5%;  Pred. No. 2.4e-168;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    6;  Gaps    1;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60
```

```
Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           ||||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy     436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy     496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy     556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTC:TGGTGATCGGCGCCGGCACC 615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy     616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy     676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy     736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy     796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy     856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy     916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy     976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy    1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy    1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy    1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy    1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1257
           |||||||||||||||||||||||||||||||||||||||||||
Db    1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1236
```

```
RESULT   9
A87607
LOCUS       A87607       1237 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9836089.
ACCESSION   A87607
VERSION     A87607.1  GI:6736247
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    386 g    221 t
ORIGIN
```

Query Match 96.8%; Score 1219.4; DB 5; Length 1237;
Best Local Similarity 99.4%; Pred. No. 3e-168;
Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

```
Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCATACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           ||||||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy     436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy     496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy     556 CTGAT:GGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
```

```
Db    535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675

Db    595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735

Db    655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795

Db    715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855

Db    775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915

Db    835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975

Db    895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035

Db    955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095

Db   1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155

Db   1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215

Db   1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCG 1258

Db   1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCG 1237
```

```
RESULT  10
A89747
LOCUS       A89747       1237 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 4 from Patent WO9832862.
ACCESSION   A89747
VERSION     A89747.1  GI:6738281
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1237)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    386 g    221 t
ORIGIN
```

Query Match 96.8%; Score 1219.4; DB 5; Length 1237;
Best Local Similarity 99.4%; Pred. No. 3e-168;

Matches 1236; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

```
Qy     16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75

Db      1 ATCTTGCAGATTAATCGAACTTTCTTCATACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy     76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135

Db     61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195

Db    115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255

Db    175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315

Db    235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy    316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375

Db    295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435

Db    355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495

Db    415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555

Db    475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615

Db    535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675

Db    595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735

Db    655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795

Db    715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855

Db    775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915

Db    835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975

Db    895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035

Db    955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095

Db   1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074
```

```
Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCG 1258
          |||||||||||||||||||||||||||||||||||||||||||
Db   1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCG 1237


RESULT  11
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN

  Query Match             96.5%;  Score 1216;  DB 5;  Length 1235;
  Best Local Similarity   99.3%;  Pred. No. 9.3e-168;
  Matches 1232;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy     16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy     76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
          |||||||||||||||||||||||||||||||||||||      ||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy    316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGC 1256
          |||||||||| |||||||||| |||||||||||||||| |
Db   1195 GCCGAGCACACNTCGGGAGTAANGGAAGCGATGATGTCGNC 1235


RESULT  12
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
```

```
ORIGIN

  Query Match             96.5%;  Score 1216;  DB 5;  Length 1235;
  Best Local Similarity   99.3%;  Pred. No. 9.3e-168;
  Matches 1232;  Conservative    0;  Mismatches    3;  Indels    6;  Gaps    1;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           |||||||||||||||||||||||||||||||||||||      ||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy     436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy     496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy     556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy     616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy     676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCC CACTCGC 714

Qy     736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTCCCGACCTGGTGATTGGGGCC 795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGT CCGACCTGGTGATTGGGGCC 774

Qy     796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy     856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy     916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy     976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGAC AACGCGACGATG 1035
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||

Db     955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy    1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy    1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy    1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy    1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGC 1256
           ||||||||||| |||||||||| |||||||||||||||| |
Db    1195 GCCGAGCACACNTCGGGAGTAANGGAAGCGATGATGTCGNC 1235


RESULT  13
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Locatic /Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN


  Query Match             96.1%;  Score 1210.4;  DB 5;  Length 1228;
  Best Local Similarity   99.4%;  Pred. No. 6e-167;
  Matches 1227;  Conservative    0;  Mismatches    1;  Indels    6;  Gaps    1;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCATACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           |||||||||||||||||||||||||||||||||||||      ||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414
```

```
Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy   1156 CTGGGGGTGCCGTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGA 1249
          |||||||||||||||||||||||||||||||||
Db   1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGA 1228
```

```
RESULT  14
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
```

```
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t
ORIGIN
```

Query Match 96.1%; Score 1210.4; DB 5; Length 1228;
Best Local Similarity 99.4%; Pred. No. 6e-167;
Matches 1227; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

```
Qy     16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
          ||||||||||||||||| ||||||||| |||||||||||||||||||||||||||||||
Db      1 ATCTTGCAGATTAATCGAACTTTCTTCATACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy     76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
          |||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy    316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    295 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 354

Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    355 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 414

Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 474

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    475 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 534

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    535 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 594

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    595 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 654

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    655 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 714

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    715 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 774

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    775 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 834

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    835 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 894

Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
```

```
Db    895 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 954

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1014

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1015 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1074

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1075 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1134

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1135 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1194

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGA 1249
          |||||||||||||||||||||||||||||||||
Db   1195 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGA 1228


RESULT  15
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             95.8%;  Score 1207;  DB 5;  Length 1235;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-166;
  Matches 1234;  Conservative    0;  Mismatches    1;  Indels    7;  Gaps    2;

Qy      16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
           |||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db      61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACG------AATTCCGGGTGGCCATC 114

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATCAGGTGCTCATCCAGGCAGGT 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 174

Qy     196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 234

Qy     256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 294

Qy     316 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||

Db     295 GCGGCGGAATACGGCCGCCTGCGACACGGGC-GATCTTGTTCACGTTCTTGCATTTGGCC 353

Qy     376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 413

Qy     436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     414 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 473

Qy     496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     474 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 533

Qy     556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     534 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 593

Qy     616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     594 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 653

Qy     676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     654 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 713

Qy     736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 773

Qy     796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     774 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 833

Qy     856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     834 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 893

Qy     916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     894 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 953

Qy     976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1013

Qy    1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1073

Qy    1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1074 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1133

Qy    1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1134 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1193

Qy    1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1257
           ||||| |||||||||||||||||||||||||||||||||||||
Db    1194 GCCGANCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCC 1235


Search completed: June 22, 2000, 14:50:07
Job time: 17422 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 10:11:14 ; Search time 458.59 Seconds
(without alignments)
687.416 Million cell updates/sec

Title: US-09-362-485-1
Perfect score: 1260
Sequence: 1 GAATTCCCATCAGCAATCTT.....AAGCGATGATGTCGGCCGCG 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1260	100.0	1260	1	V49510	Mycobacterium sp.
2		1260	100.0	1260	1	V49625	Mycobacterium tube
3		1245	98.8	1245	1	V49626	Mycobacterium tube
4		467.6	37.1	682	1	V49511	Mycobacterium mari
5		261.2	20.7	1125	1	N91423	Sequence of heat-r
c 6		198.4	15.7	28171	1	V52155	Streptococcus pneu
c 7		178.4	14.2	9280	1	V74442	Staphylococcus aur
8		94.8	7.5	1074	1	T67971	H. pylori membrane
c 9		69.2	5.5	544	1	V43039	Streptococcus pneu
10		57.4	4.6	31096	1	V74370	Staphylococcus aur
c 11		50.8	4.0	390	1	Q21833	Randomising oligon
c 12		50.8	4.0	390	1	Q36859	PCR primer for 5'
13		50.4	4.0	15872	1	T68715	Streptomyces venez
14		49	3.9	1269	1	Q03065	Sequence encoding
15		47.6	3.8	1413	1	X34205	Mycobacterium spec
16		47.6	3.8	1722	1	X34206	Mycobacterium spec
c 17		47.6	3.8	3946	1	T93610	Mycobacterium tube
18		47.6	3.8	28598	1	T06769	Sorangium cellulos
19		47.6	3.8	28958	1	T89956	Sorangium cellulos
20		47.6	3.8	49377	1	V05287	The soraphen biosy
c 21		46.4	3.7	985	1	V44439	Mycobacterium tube
c 22		46.4	3.7	985	1	V64548	M. tuberculosis im
23		46	3.7	1620	1	Q22482	groEL-1 gene codin
24		46	3.7	2582	1	T73117	Actinoplanes sp. a
25		46	3.7	2668	1	Q22485	groEL-1 gene. Reco
26		46	3.7	17955	1	V56642	Actinoplanes sp. a
27		45.4	3.6	833	1	Q64203	snaB gene encoding
28		45.4	3.6	5392	1	Q64201	Sequence comprisin
29		42.4	3.4	1833	1	Q64206	snbR gene encoding
30		42.4	3.4	2185	1	V84066	Clone P5-4 encodin
31		42.4	3.4	2186	1	Q99365	S. lividans protea
32		42.4	3.4	24379	1	T93095	Streptomyces freno
33		42.4	3.4	24379	1	V25925	Streptomyces roseo
34		42.2	3.3	329	1	V44425	Mycobacterium tube
35		42.2	3.3	3.9	1	V64534	M. tuberculosis im
36		42	3.3	3946	1	T93610	Mycobacterium tube
c 37		42	3.3	4018	1	Q63879	Polyhydroxyalkanoa
38		41.8	3.3	882	1	V44403	Mycobacterium tube
39		41.8	3.3	882	1	V64512	M. tuberculosis im
40		41.6	3.3	12036	1	Q04668	FHA structural gen
41		41.2	3.3	2404	1	Q05926	Sequence encoding
42		41	3.3	29879	1	Q46806	eryA region of S.
c 43		40.8	3.2	2151	1	Q76252	HSV-2 protease, IC
c 44		40.8	3.2	2151	1	Q76261	HSV-2 protease/ICP
c 45		40.8	3.2	2472	1	Q84671	HSV-2 UL26 gene. N

ALIGNMENTS

RESULT 1
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 C; 225 T;

Query Match 100.0%; Score 1260; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1.3e-256;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60

Qy      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120

Qy     121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180

Qy     181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240

Qy     241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
```

```
Qy    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360

Qy    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420

Qy    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480

Qy    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540

Qy    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600

Qy    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660

Qy    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720

Qy    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780

Qy    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840

Qy    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900

Qy    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960

Qy    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020

Qy   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080

Qy   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140

Qy   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200

Qy   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAG GAAGCGATGATGTCGGCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
```

RESULT 2
V49625
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.

PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 1260; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1.3e-256;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAATTCCCATCAGCAATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTA 60

Qy     61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCGAGAGGGGTAATCATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAA 120

Qy    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTG 180

Qy    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCA 240

Qy    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAG 300

Qy    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACG 360

Qy    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACG 420

Qy    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATG 480

Qy    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAA 540

Qy    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTG 600
```

```
Qy    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACC 660

Qy    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGC 720

Qy    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGAC 780

Qy    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCA 840

Qy    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGC 900

Qy    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGTTTCGAAGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACG 960

Qy    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTG 1020

Qy   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGC 1080

Qy   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAA 1140

Qy   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGGGTGGCCACCGACCTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTC 1200

Qy   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TCGGCCGCTCGTTACGCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
```

RESULT 3
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 98.8%; Score 1245; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 1.8e-253;
Matches 1245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     16 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 75
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATC 60

Qy     76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 120

Qy    136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGT 180

Qy    196 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGTC 240

Qy    256 GGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 315
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCAC 3CCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATA 300

Qy    316 GCGGC GAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCC 360

Qy    376 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAG 420

Qy    436 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGT 480

Qy    496 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTG 540

Qy    556 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACC 600

Qy    616 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGAC 660

Qy    676 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGC 720

Qy    736 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCC 780

Qy    796 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATG 840

Qy    856 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCA 900
```

```
Qy    916 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGACCGACCACCTACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTG 960

Qy    976 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATG 1020

Qy   1036 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAATCCGGCA 1080

Qy   1096 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCCACCGAC 1140

Qy   1156 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGGGGGTGCCGTTCACCGAGCCCGCCAGCGTGCTGGCCTGACTCTCGGCCGCTCGTTAC 1200

Qy   1216 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1260
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GCCGAGCACACGTCGGGAGTAAGGGAAGCGATGATGTCGGCCGCG 1245
```

RESULT 4
V49511
ID V49511 standard; DNA; 682 BP.
AC V49511;
DT 20-OCT-1998 (first entry)
DE Mycobacterium marinum Mar3 DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium murinum.
FH Key Location/Qualifiers
FT CDS 1. .682
FT /*tag= a
FT /codon_start= 3
FT /product= "Mar3"
FT /note= "Alanine dehydrogenase"
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
DR P-PSDB; W64481.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Claim 1; Page 34-35; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein, Mar3
CC isolated from Mycobacterium murinum. This protein is used to diagnose
CC tuberculosis and other mycobacterial infections (including 'swimmers'
CC disease', caused by M. marinum, a fish pathogen) in humans or animals.
CC The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 682 BP; 105 A; 254 C; 225 G; 98 T;

Query Match 37.1%; Score 467.6; DB 1; Length 682;
Best Local Similarity 80.4%; Pred. No. 3.8e-90;
Matches 548; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```
Qy    119 AATTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGG 178
          |||||||| | || ||||||||||||||||||||| |   | |||   || ||||| ||||
Db      1 AATTCCGGTTAGCGATCACCCCGGCCGGCGTCGCCGCCTTGACCAAGCGCGGCCACGAGG 60

Qy    179 TGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGG 238
          |||| |||||||| |||||||||| ||||| || ||| ||||||| || ||||||||| |
Db     61 TGCTGATCCAGGCCGGTGCCGGAGAAGGCTCCGCCATCTCCGACGCCGACTTCAAGGCCG 120

Qy    239 CAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCA 298
          | || || || ||| || ||||||||||||||||||||||| || || ||  | ||||||
Db    121 CCGGTGCCCAGCTGATCAGCACCGCCGACCAGGTGTGGGCGGATGCGGACCTGCTGCTCA 180

Qy    299 AGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCA 358
          ||||||| |||||||| | | | || |||||||| ||||| |  || |||| | ||||||
Db    181 AGGTCAAGGAACCGATCGAGTCCGAGTACGGCCGGCTGCGCCGGGGCCAGACCCTGTTCA 240

Qy    359 CGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCA 418
          | | | ||||  ||||||| || ||  | ||||||||||  || || | ||||||||||
Db    241 CCTACCTGCACCTGGCCGCCTCGCGCCCCTGCACCGATGCCCTGCTGAAGTCCGGCACCA 300

Qy    419 CGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGA 478
          |||| || |||||||||| || ||||||||||||||||||| | || ||||| ||||| |
Db    301 CGTCCATCGCCTACGAGACGGTGCAGACCGCCGACGGCGCATTGCCGCTGCTGGCCCCCA 360

Qy    479 TGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCC 538
          ||||||| |||||||| || ||  ||||||| | ||| || ||||||||||||||| ||||
Db    361 TGAGCGAGGTCGCCGGGCGCCTGTCCGCCCAAGCTGGGGCCTACCACCTGATGCGCACCC 420

Qy    539 AAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGG 598
          | || || ||||| |||||||||||||| || ||||||||| | || ||||||||||||||
Db    421 ACGGCGGTCGCGGCGTGCTGATGGGCGGCGTCCCCGGCGTCAAGCCTGCCGACGTCGTGG 480

Qy    599 TGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGA 658
          |||||||||| ||||| ||||| |||||||| |||||| ||||||||||||||||||||||
Db    481 TGATCGGCGCGGGCACGGCCGGATACAACGCCGCCCGCGTCGCCAACGGCATGGGCGCGA 540

Qy    659 CCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCG 718
            || || || || ||  |||||||| |||| || || ||  ||||||| |||||| |||
Db    541 TGGTCACCGTGCTGGATGTCAACATCAACAAGCTCCGCCAGATCGACGCGGAGTTCGGCG 600

Qy    719 GCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCG 778
          | ||  |||  || |||||||| ||| ||  ||| ||||||| ||| |     | ||||
Db    601 GTCGCGTCCGGACCCGCTACTCGTCGACCCTCGACCTCGAGGATGCGGCAGTCCACGCCG 660

Qy    779 ACCTGGTGATTGGGGCCGTCCT 800
          || ||||||| ||||||||||| 
Db    661 ACATGGTGATCGGGGCCGTCCT 682
```

RESULT 5
N91423
ID N91423 standard; DNA; 1125 BP.
AC N91423;
DT 01-FEB-1991 (first entry)
DE Sequence of heat-resistant alanine dehydrogenase (AH) gene with mol. wt.
DE 2 MD or less
KW Enzyme; ds.
OS Bacillus stearothermophilus IFO 12550.
PN J01043194-A.
PD 15-FEB-1989.
PF 10-AUG-1987; 200524.
PR 10-AUG-1987; JP-200524.
PA (NIRA) Unitika Kk.
DR WPI; 89-096096/13.
PT Recombinant plasmid for transforming Escherichia coli -
PT obtd. by connecting heat resistant alanine dehydrogenase gene to
PT vector plasmid
PS Disclosure; Fig 2, p693 ; 9pp; Japanese.
CC A recombinant plasmid contg. heat-resistant alanine dehydrogenase (AH)
CC gene with mol. wt. 2 or less MD is claimed. Cells transformed with the
CC vector produce high levels of heat-resistant AH. Transformed E. coli
CC is useful as a clinical inspection reagent.
SQ Sequence 1125 BP; 246 A; 295 C; 367 G; 217 T;

Query Match 20.7%; Score 261.2; DB 1; Length 1125;
Best Local Similarity 55.9%; Pred. No. 9.2e-47;

Matches 604; Conservative 0; Mismatches 458; Indels 18; Gaps 5;

```
Qy      76 ATGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCAATTCCGGGTGGCCATC 135
Db       1 ATGAAGATCGGCATTCCAAAAGAAATCAAAAACAATG------AAAACCGCGTCGCCATC 54

Qy     136 ACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCT---CATCCAGGCA 192
Db      55 ACTCCGGCAGGCGTGATGACGCTCGTCAAAGCGGGGCATGAGGTGGTGTATGTGGAGACG 114

Qy     193 GGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTG 252
Db     115 GAAGGCGGCGCTGGGTCGGGGTTTTCCGATTCCGAGTATGAAAAAGCCGGGGCAGCTGAT 174

Qy     253 GTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCG 312
Db     175 CGTTGCCGAACGTGGAGAGATGCTTGGACGGCGGAGATGGTGTTGAAAGTGAAAGAGCCG 234

Qy     313 ATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTG 372
Db     235 CTGGCTCGAGAGTTCCGCTATTTTCGCCCCGGATTGATTTTGTTTACGTATTTGCATTTA 294

Qy     373 GCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTAC 432
Db     295 GCCGCGGCCGAACGCGTCACGAAAGCCGTCGTCGAGCAAAAAGTGGTCGGCATCGCTTAC 354

Qy     433 GAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCC 492
Db     355 GAGACGGTGCAGCTGGCGAACGGCTCGCTGCCACTGTTGACGCCGATGAGTGAAGTCGCC 414

Qy     493 GGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGT 552
Db     415 GGCCGCATGTCGGTGCAAGTCGGCGCCCAGTTTCTCGAGAAGCCGCACGGGGGAAAGGGC 474

Qy     553 GTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGC 612
Db     475 ATTTTGCTTGGCGGCGTGCCCGGAGTGCGGCGCGGCAAAGTGACGATCATCGGCGGCGGA 534

Qy     613 ACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTA 672
Db     535 ACGGCGGGGACGAACGCGGCGAAAATCGGGTCGGTCTCCGGTCAGACGCTGACGATTTTG 594

Qy     673 GACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACT 732
Db     595 GACATTAACGCCGAGCGGCTGCGCGAGCTCGATGATTTGTTCGGCGACCACGTGACGACG 654

Qy     733 CGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGG 792
Db     655 CTCATGTCCAACTCGTACCATATCGCCGAGTGCGTGCGCGAATCGGATTTGGTCGTCGGT 714

Qy     793 GCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCAT 852
Db     715 GCCGTCTTGATCCCGGGGGCGAA---AGCGAAGCTGGTGACGGAAGAGATGGTGCGCTCG 771

Qy     853 ATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGACCAGGGCGGCTGT.TCGAA--- 910
Db     772 ATGACGCCGGGATCGGTGTTGGTCGTAATCGCCATTGACCAAGGCGGCATTTTCGAAACG 831

Qy     910 GGCTCACGACCGACCACCTACGACCACCCGACGTTCGC GTGCACGACACGCTGTTTTAC 969
Db     832 ACCGACCGACCGTACGACGACGACGATCCGACATACGTCAAGCACGGCGTCGTCCATTAC 891

Qy     970 TGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCG 1029
Db     892 GCCGTCGCCAACATTGCCGGC---GTGCCGCGCACGTCGACATTCGCGCTTACGAACGTC 948

Qy    1030 ACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGGCGGGCGGCGTGCCGGTCGAAT 1089
Db     949 ACGATCCCATACGCCTTGCAAATCGCCAACAAAGGCTACCGCGCCGGTTGCTTGGATAAC 1008

Qy    1090 CCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGGGGCGTTACTGTCCGAACGGGTGGCC 1149
Db    1009 CCGGCGCTGTTAAAAGGGATCAACACGCTCGACGGGCACATCGTGTACGAAGCGGTCGGC 1068
```

RESULT 6
V52155/c
ID V52155 standard; DNA; 28171 BP.
AC V52155;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:22.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 273-289; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 28171 BP; 8507 A; 5080 C; 6440 G; 8144 T;

Query Match 15.7%; Score 198.4; DB 1; Length 28171;
Best Local Similarity 53.4%; Pred. No. 1.9e-33;
Matches 516; Conservative 0; Mismatches 431; Indels 19; Gaps 4;

```
Qy     104 AAAACAACGAATTCCAATTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCC 163
Db   21689 AAATTAAAAATAACGAAAACCGTGTCGCCCTCACACCTGCAGGTGTTCATAGCTTAGTTA 21630

Qy     164 GTCGTGGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACG 223
Db   21629 GTCGTGGTCATCGTGTCCTTATCGAAACAAATGCTGGTCTCGGTTCTGGCTTTACTGATG 21570

Qy     224 CGGATTTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACG 283
Db   21569 CTGACTATCAAAAGCAAGGAGCTGAGATTGTCGCTACTGCTGGTGAAGCTTGGGC---AG 21513

Qy     284 CTGATTTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACG 343
Db   21512 CAGAGTTGGTTGTGAAAGTAAAAGAATCTTTAAGTTCTGAATACGGTTACTTGCGCGACG 21453

Qy     344 GGCAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGT 403
Db   21452 ATCTTCTTCTCTTCACCTACTTGCACATGGCCGCTGCTCCAGAATTAGCAGATGCTATGT 21393
```

```
Qy     404 TGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTAC 463
           |             ||  |||   |    || || || |    |    | ||   |||||
Db   21392 T------------AACAGCAAAAACAACTGAAACTGTTCGTGACAATCAAGGACAACTAC 21345

Qy     464 CCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACC 523
           | || || |  || ||||| || || || ||||| | || |  ||  | || ||| ||
Db   21344 CGCTCCTCGTTCCTATGAGTGAGGTTGCAGGTCGTATGGCTGTTCAAATCGGAGCTCACT 21285

Qy     524 ACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAAC 583
           |||  |            || |||   ||||| ||  | || || || || || || ||   |
Db   21284 TCCTTACTAAGCAAGCTGGTGGCTCTGGTGTTCTACTTGGTGGTGTACCAGGTGTTCCAA 21225

Qy     584 CGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCA 643
           |   | ||     | ||||| |  ||   || |||      | || ||||||||||||
Db   21224 AAGGAAAAGTAACTATCATCGGTGGTGGTGTCGTCGGTACACATGCTGCCCGCATCGCCC 21165

Qy     644 ACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCG 703
           ||  | || ||    || ||  || |||| || |    | |    ||          || |
Db   21164 TTGGTCTTGGTGCTCAAGTGACTATTTTAGATATTAGTTCCAAGCGTCTCTCAGTTCTAG 21105

Qy     704 ACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCA CGGCCTACGAGCTCG GGGTG 763
           | |  |  ||  |  | |  || || ||||     ||     | | | |  | || |
Db   21104 AAGAAGTCTTTGGAAGTCAAATTCAAACTCTTATGTCTAATTCATTCAACATTGAAGCAA 21045

Qy     764 CCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCA 823
           || | |  ||| ||  ||||||||||| ||| | ||  | || || || || |||||
Db   21044 GTGTGAGAGATGCTGATGTGGTGATTGGAGCCATTCTCATCCCTGGTGCAAAAGCACCGG 20985

Qy     824 AATTAGTCTCGAATTCACTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAG 883
           ||||  ||  |  ||    | |||   || |||   |||||  | ||| | || ||  | |
Db   20984 AATTGGTGACAGATGAGATGGTCAAACAAATGCGTCCAGGCTCTGTA-TCGTTGACGTTG 20926

Qy     884 CCATCGACCAGGGCGGCTGTTTCGA---AGGCTCACGACCGACCACCTACGACCACCCGA 940
           |  | |||||  || |||  | ||||   ||     ||   ||| ||  ||||  | ||
Db   20925 CTGTTGACCAAGGTGGCGTTATCGAAACAGCTGACCGTGTGACAACGCACGATGAACCCG 20866

Qy     941 CGTTCGCCGTGCACGACACGCTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGA 1000
           |  |     ||||     ||     ||   ||| || || || || |   ||||  |
Db   20865 TCTATGAAAAACACGGTGTTCTCCACTATGCCGTTGCCAATATCCCTGGTGCGGTTGCTC 20806

Qy    1001 AGACGTCGACCTACGCGCTGACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACC 1060
           || || |||  ||| || ||||| |  ||  | || ||| |    |   | || | |
Db   20805 GCACTTCAACCATCGCCCTAACCAATGTCACTCTTCCTTATATCGAAGCTTTGGCTGGCA 20746

Qy    1061 ATGGCT 1066
           | || |
Db   20745 AAGGAT 20740
```

```
RESULT   7
V74442/c
ID   V74442 standard; DNA; 9280 BP.
AC   V74442;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #131.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    841. .900
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    2641. .2700
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT
FT   misc_feature    4441. .4500
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    6241. .6300
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    8041. .8100
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 710-715; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   9280 BP;    2958 A;    1649 C;    1309 G;    3060 T;
```

```
  Query Match             14.2%;  Score 178.4;  DB 1;  Length 9280;
  Best Local Similarity   50.6%;  Pred. No. 2.8e-29;
  Matches  484;  Conservative    0;  Mismatches  466;  Indels    6;  Gaps    2;

Qy     172 CATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTC 231
           |||   ||  |  |  |  ||  ||| ||    || ||      |    || |  |||| |
Db    9280 CATACTGTTTTAGTGGAAACAAATGCGGGTTCAGGATCATTCTTTGAAGATGTAGATTAC 9221

Qy     232 AAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTA 291
           || |  ||||| ||  |  | || |      |   | || |||   || |  ||| |
Db    9220 AAAGAAGCAGGTGCTGAGATTGTTGCTGAACAAGCAAAAGTTTGG---GATGTGGATATG 9164

Qy     292 TTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATC 351
           |  | || || ||||||||  | |   | |||||       |   | | ||||   |
Db    9163 GTTATTAAAGTTAAAGAACCACTTGAATCTGAATATCCATATTTTAAAGAAGGGCTTGTA 9104

Qy     352 TTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCC 411
           || ||||| |   | ||||| ||       |       |  ||  | || ||| | |||
Db    9103 TTATTCACTTATCTTCATTTAGCAAATGAAGAAAAATTAACACAAGCTTTGATAGATAGA 9044

Qy     412 GGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTT 471
                |     ||||| || ||||| || |||    | ||| |  |  ||||  || |
```

```
Db    9043 AAAGTAATTAGTATTGCATATGAGACTGTGCAGTTACCAGACCGATCTTTACCATTGTTA 8984

Qy     472 GCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATG 531

Db    8983 TCACCAATGAGTGAGGTAGCAGGAAGAATGTCAGCTCAAGTTGGCGCAGAGTTCCTACAA 8924

Qy     532 CGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGAC 591

Db    8923 AAACTTAATGGTGGTATGGGAATTCTACTAGGTGGTGTCCCTGAGTACCTAAGGGTAAA 8864

Qy     592 GTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATG 651

Db    8863 GTAACTATTATCGGTGGTGGTCAAGCAGGAACAAATGCAGCTAAAATTGCACTAGGACTA 8804

Qy     652 GGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAG 711

Db    8803 GGTGCAGATGTTACGATTTTAGATGTTAATCCAAAGCGTTTACAACAATTAGATGATTTA 8744

Qy     712 TTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAA 771

Db    8743 TTCGGTGGACGTGTACATACAATTATGTCAAATCCGTTGAATATTGAGTTGTATGTTAAA 8684

Qy     772 CGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTC 831

Db    8683 CAAAGTGATTTAGTAATAGGTGCAGTTTTAATTCCAGGTGCTAAAGCGCCAAGACTTGTA 8624

Qy     832 TCGAATTCACTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGAC 891

Db    8623 ACAGAAGACATGATTAAACAAATGAAAAATGGGTCAGTTATTATTGACATTGCTATTGAT 8564

Qy     892 CAGGGCGGCTGTTTCGAAGGCTC---ACGACCGACCACCTACGACCACCCGACGTTCGCC 948

Db    8563 CAAGGCGGTATTTTTGAAACAACTGATAAAATTACGACACATGATGATCCTACATATATT 8504

Qy     949 GTGCACGACACGCTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCG 1008

Db    8503 AAGCATGGTGTGGTTCATTATGCAGTTGCAAATATGCCAGGTGCAGTACCGCGTACTTCG 8444

Qy    1009 ACCTACGCGCTGACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCTGG 1068

Db    8443 ACGTTAGCTTTAAATAATGCTACGCTACCTTATGCGCTCATGCTAGCTAATAAAGGGTAT 8384

Qy    1069 CGGGCGGCGTGCCGGTCGAATCCGGCACTAGCCAAAGGTCTTTCGACGCACGAAGG 1124

Db    8383 AGAGAAGCATTTAAATCAAATCAACCATTATCATTAGGTTTAAATACTTACAAAGG 8328
```

```
RESULT   8
T67971
ID   T67971 standard; DNA; 1074 BP.
AC   T67971;
DT   15-JUL-1997  (first entry)
DE   H. pylori membrane protein ORF 05cp20518orf61.
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   membrane; amino acid; metabolism; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1. .1074
FT                   /*tag= a
FT                   /note= "no stop codon given"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20718.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 27; Page 819; 1481pp; English.
CC   The present sequence encodes a Helicobacter pylori membrane
CC   protein likely to contain four membrane spanning regions.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the
CC   bacterial DNA. The sequences were analysed for ORF of at least 180
CC   nucleotides, and the predicted coding regions defined by computer
CC   evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences
CC   of interest, particular regions can be isolated from H. pylori by
CC   PCR amplification for recombinant polypeptide production, e.g. in
CC   E. coli hosts.
SQ   Sequence   1074 BP;    302 A;    189 C;    297 G;    286 T;
```

```
  Query Match             7.5%;  Score 94.8;  DB 1;  Length 1074;
  Best Local Similarity   46.1%;  Pred. No. 8.4e-12;
  Matches  370;  Conservative    0;  Mismatches  417;  Indels   15;  Gaps    1;

Qy     286 GATTTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGG 345

Db     133 GATTTGGTGGTCAAATGCAAAGAGCCTTTAGAGCATGAATACCCTTTGCTCAAAGAAAAA 192

Qy     346 CAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTG 405

Db     193 GCGACTCTGTTTAGTTATTTGGATTTAGCGTATCAAAAAAGCTTGTGTGAAATGTTTATT 252

Qy     406 GATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCC 465

Db     253 AATAAAAAAATCACTTCTATTTGCACTGAAACCATTGCCGGGCCTAAAAACGACTACCCT 312

Qy     466 CTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGC--------- 517

Db     313 ATTTTAGCGCCTATGAGCGTGGTGGCTGGGAGGTTGGCTGCGCATTTAGTCCAGCATTAT 372

Qy     517 ------GCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTG 570

Db     373 TTACTGGCTTTAGAGCATGTTAAAGGGTTTATGGGTAAGGGGGTCATGCTAGGGGGTTTG 432

Qy     571 CCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCA 630

Db     433 TCGGGTGCGCAAAGGGCTAAAATCGTCGTAATTGGAGGCGGTGTGGTTGGCATGGAGAGC 492

Qy     631 GCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAA 690

Db     493 GCGAAAGTCTTAAGCCAAATGGGGGCTAAAGTAACGATTTTAGAATTAGACTACGCTAAA 552

Qy     691 CTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTAC 750

Db     553 TTACAAAACCACCCTTATTATCATTTGTATGATTTAGAAGTCTTAAGCGTGAATGAAGCC 612

Qy     751 GAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGC 810

Db     613 AATATCATTCAAGCCTTAAACGGGGCGGTGGGGCTAGTGGGAGCGGTGCTGGTTACAGCG 672

Qy     811 GCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATGAAACCAGGTGCGGTA 870

Db     673 AGCCAAACCCCTAAAGTGATCTTAAGAAGCATTTAAAATACATGCAGACACAAGGGGTA 732

Qy     871 CTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCACGACCGACCACCTAC 930

Db     733 GTCATTGATGTGGCTTGCAGTTTAGGGGGGTGCATAAGAGCCATACGCCAAGCAAGCCAT 792

Qy     931 GACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTGGCGAACATGCCCGCC 990

Db     793 TCTAACCCGGTGTATGTGGAAGAAAGTTTGTTGCATTATGGCGTGCCGAACATGCCAGGG 852
```

```
Qy    991 TCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATGCCGTATGTGCTCGAG 1050
          ||  | || |||    |    ||    | | | ||||   |||||||| || |  |
Db    853 ATTGTCGCTAAAACGAGCTCTACGGCTTATAGCCATGCGAGTGTGCCGTATTTGTTGTAT 912

Qy   1051 CTTGCCGACCATGGCTGGCGGG 1072
            |   || |||||||  |  ||
Db    913 TATTTAGAGCATGGCTTGAAGG 934


RESULT   9
V43039/c
ID   V43039 standard; DNA; 544 BP.
AC   V43039;
DT   09-NOV-1998  (first entry)
DE   Streptococcus pneumoniae polypeptide coding region.
KW   Polypeptide; ORF; open reading frame; infection; bacterial;
KW   streptococcal; bacteremia; diagnosis; prophylaxis; ds.
OS   Streptococcus pneumoniae.
FH   Key             Location/Qualifiers
FT   CDS             complement (59. .334)
FT                   /*tag= a
FT                   /note= "polypeptide"
PN   WO9823631-A1.
PD   04-JUN-1998.
PF   24-NOV-1997; U21976.
PR   27-NOV-1996; US-031879.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI   Reid RH, Zarfos PN;
DR   WPI; 98-322654/28.
DR   P-PSDB; W62769.
PT   Streptococcus pneumoniae polynucleotides - useful for developing
PT   products for diagnosis, prevention and treatment of infections e.g.
PT   pneumonia, bacteremia, meningitis or endocarditis
PS   Claim 1; Page 165-166; 181pp; English.
CC   The sequence is that of a Streptococcal polypeptide coding region.
CC   The polypeptide can potentially be used for the diagnosis and
CC   prevention of bacterial infections, especially SP infection.
CC   It may be used for the treatment of diseases such as otitis media,
CC   conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC   empyema, endocarditis or infection of the cerebrospinal fluid.
SQ   Sequence   544 BP;    149 A;    121 C;    124 G;    150 T;


  Query Match             5.5%;  Score 69.2;  DB 1;  Length 544;
  Best Local Similarity   53.4%;  Pred. No. 1.9e-06;
  Matches  191;  Conservative    0;  Mismatches  163;  Indels    4;  Gaps    2;

Qy    712 TTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAA 771
          ||  |  | |  || || | ||     ||     | | | |  | || |      | | |
Db    543 TTTGGAAGTCAAATTCAAATTCTTATATCTAATTCATTCAACATTGAAGCAAGTATGAGA 484

Qy    772 CGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAAATTAGTC 831
            ||| ||  ||||||||||| ||| | ||  | || || || || |||||  |||| ||
Db    483 GATGCTGATGTGGTGATTGGAGCCATTCTCATCCCTGGTGCAAAAGCACCGGAATTGGTG 424

Qy    832 TCGAATTCACTTGTCGCGCATATGAAACCAGGTGCGGTACTGGTGGATATAGCCATCGAC 891
          |  ||    | |||   || |||   |||||  | ||| | || ||  | ||  | |||
Db    423 ACAGATGAGATGGTCAAACAAATGCGTCCAGGCTCTGTA-TCGTTGACGTTGCTGTTGAC 365

Qy    892 CAGGGCGGCTGTTTCGA---AGGCTCACGACCGACCACCTACGACCACCCGACGTTCGCC 948
          || || |||  | ||||   ||     ||   ||| ||  ||||  | ||    |  |
Db    364 CAAGGTGGCGTTATCGAAACAGCTGACCGTGTGACAACGCACGATGAACCCGTCTATGAA 305

Qy    949 GTGCACGACACGCTGTTTTACTGCGTGGCGAACATGCCCGCCTCGGTGCCGAAGACGTCG 1008
          ||||     ||    ||   ||| || || || || |   ||||  |    || ||
Db    304 AAACACGGTGTTCTCCACTATGCCGTTGCCAATATCCCTGGTGCGGTTGCCCGCACTTCA 245

Qy   1009 ACCTACGCGCTGACCAACGCGACGATGCCGTATGTGCTCGAGCTTGCCGACCATGGCT 1066
          |||  ||| || ||||| |  ||  | || ||| |    |   | || | | | || |
Db    244 ACCATCGCCCTAACCAATGTCACTCTTCCTTATATCGAAGCTTTGGCTGGCAAAGGAT 187


RESULT  10
V74370
ID   V74370 standard; DNA; 31096 BP.
AC   V74370;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #59.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1201. .1260
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    3001. .3060
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    4801. .4860
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    6601. .6660
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    8401. .8460
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    10201. .10260
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    12001. .12060
FT                   /*tag= g
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    13801. .13860
FT                   /*tag= h
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    15601. .15660
FT                   /*tag= i
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
FT   misc_feature    17401. .17460
FT                   /*tag= j
FT                   /note= "these bases represent a line of missing text in
FT                    the sequence listing in the specification. They
FT                    are included to maintain the nucleotide numbering
FT                    given in the specification for this DNA sequence"
```

```
FT   misc_feature    19201. .19260
FT                   /*tag= k
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    21001. .21060
FT                   /*tag= l
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    22801. .22860
FT                   /*tag= m
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    24601. .24660
FT                   /*tag= n
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    26401. .26460
FT                   /*tag= o
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    28201. .28260
FT                   /*tag= p
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    30001. .30060
FT                   /*tag= q
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 452-469; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   31096 BP;    11857 A;    5243 C;    5477 G;    7488 T;
```

```
 Query Match             4.6%;  Score 57.4;  DB 1;  Length 31096;
 Best Local Similarity   44.8%;  Pred. No. 0.00082;
 Matches  358;  Conservative    0;  Mismatches  426;  Indels   15;  Gaps    3;

Qy    280 GACGCTGATTTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGA 339
          || |||||| |  |  |||| || |||||||| |     |||||          |  |
Db    185 GAAGCTGATCTTGTTATCAAAGTAAAAGAACCTCATGAAAGCGAATATCAATATTTCAAA 244

Qy    340 CACGGGCAGATCTTGTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCG 399
          |    || ||  | |      || || ||| | ||  | |||   |         ||
Db    245 AAGAATCAAATTATCTGGGGATTTTTACATCTAGCATCTTCAAAAGAAATAGTAGAAAAA 304

Qy    400 TTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCA 459
           ||   ||    ||    ||  | |||          || ||| |   |   | , | ||  |
Db    305 ATGCAAGAAGTTGGTGTAACTGCGATTAGTGGTGAAACCAT---TATAAAAAATGGAAAA 361

Qy    460 CTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCT 519
            |    |  | || || ||||| |   | || |||| || | | ||    | || |||
Db    362 GCAGAATTATTAGCGCCAATGAGTGCTATAGCAGGTCAACGCTCAGCAATTATGGGAGCT 421

Qy    520 TACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTC 579
          ||| ||        |   || || || |  |||    |  ||   || || | |
Db    422 TACTACTCTGAAGCACAACATGGTGGTCAAGGTACTTTAGTGACTGGTGTACATGAAAAT 481

Qy    580 GAACCGGCCGACG---------TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCA 630
          |       |             ||| |  |||| |  ||    || | |  || |||
Db    482 GTGGATATACCTGGTAGTACATATGTGATTTTCGGTGGTGGAGTAGCAGCAACAAATGCA 541

Qy    631 GCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAA 690
          ||     | |||  ||  |    || |  || |   || | ||  | |||   |||
Db    542 GCAAATGTTGCCTTGGGACTAAATGCTAAAGTAATCATTATCGAGTTAAACGATGACCGC 601

Qy    691 CTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTC---GCTACTCATCGGCC 747
          ||    | || || |    ||      |          |||      | | ||| |  |
Db    602 ATTAAATATCTTGAAGATATGTATGCAGAAAAAGATGTCACAGTAGTCAAATCAACACCA 661

Qy    748 TACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCA 807
          |  |  | |      | ||   || || |  | |||   |  | |  |  | ||
Db    662 GAAAATTTAGCAGAACAAATTAAGAAAGCAGATGTATTTATTTCTACAATTTTAATTTCA 721

Qy    808 GGCGCCAAGGCACCCAAATTAGTCTCGAATTCACTTGTCGCGCATATGAAACCAGGTGCG 867
          || || ||  | || |||||  ||  |   |    | ||       ||||||  |||| |
Db    722 GGTGCGAAACCGCCAAAATTGGTTACTCGTGAGATGGTTAAATCAATGAAAAAAGGTTCA 781

Qy    868 GTACTGGTGGATATAGCCATCGACCAGGGCGGCTGTTTCGAAGGCTCACGACCGACCACC 927
          ||| |  | |||||||| || ||||| || ||   | | |||       |||| || ||
Db    782 GTATTAATCGATATAGCTATTGACCAAGGTGGAACTATTGAAACAATTAGACCAACTACA 841

Qy    928 TACGACCACCCGACGTTCGCCGTGCACGACACGCTGTTTTACTGCGTGGCGAACATGCCC 987
              | ||   ||  |  |   | |    | |   |||  | ||  | ||     ||
Db    842 ATTTCTGATCCAGTGTATGAAGAAGAAGGTGTGATTCATTATGGTGTACCAAATCAACCA 901

Qy    988 GCCTCGGTGCCGAAGACGTCGACCTACGCGCTGACCAACGCGACGATGCCGTATGTGCTC 1047
          |   | || || |  || || ||    ||  |  |  |  |  |  ||    ||| |  |
Db    902 GGAGCAGTCCCAAGAACTTCAACAATGGCATTAGCACAAGGAAATATTGATTATATATTA 961

Qy   1048 GAGCTTGCCGACCATGGCT 1066
          ||  ||   ||| | ||||
Db    962 GAAATTTGTGACAAAGGCT 980
```

```
RESULT  11
Q21833/c
ID   Q21833 standard; DNA; 390 BP.
AC   Q21833;
DT   08-JUN-1992  (first entry)
DE   Randomising oligonucleotide used in SPERT mRNA prepn.
KW   Systematic polypeptide evolution by reverse translation; SPERT;
KW   ligand binding; ss.
OS   Synthetic.
PN   WO9202536-A.
```

PD 20-FEB-1992.
PF 01-AUG-1991; U05463.
PR 02-AUG-1990; US-561968.
PA (COLS) UNIV OF COLORADO.
PI Gold L, Tuerk C;
DR WPI; 92-080018/10.
PT New method of systematic polypeptide evolution by reverse
PT translation - by linking each polypeptide in sample mixt. to
PT individualised mRNA allowing further synthesis of selected
PT polypeptide(s)
PS Example; Page 55; 102pp; English.
CC The sequence is that of an example randomising oligonucleotide which
CC is used in the prepn. of mRNA encoding candidate polypeptides for the
CC method of systematic polypeptide evolution by reverse translation
CC (SPERT). The method provides a rapid way of isolating and identifying
CC polypeptide ligands which bind to target mols. The polypeptide ligands
CC can be used in e.g. assay methods, diagnostic procedures, cell sorting,
CC as activators or inhibitors of target mol. function, as probes, as
CC sequestering agents, drug delivery vehicles, modifiers of hormone
CC action and as catalysts. See also Q21830-Q21832.
SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T;

Query Match 4.0%; Score 50.8; DB 1; Length 390;
Best Local Similarity 46.2%; Pred. No. 0.014;
Matches 169; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```
Qy    447 CGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGC 506
Db    366 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 307

Qy    507 CCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGG 566
Db    306 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 247

Qy    567 GGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAA 626
Db    246 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 187

Qy    627 CGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGA 686
Db    186 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 127

Qy    687 CAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGC 746
Db    126 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 67

Qy    747 CTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCC 806
Db     66 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCCA 7

Qy    807 AGGCGC 812
Db      6 TGGCCC 1
```

RESULT 12
Q36859/c
ID Q36859 standard; DNA; 390 BP.
AC Q36859;
DT 22-JUN-1993 (first entry)
DE PCR primer for 5' fixed sequence contg. T7 promoter and RBS.
KW Systematic peptide evolution by reverse translation; SPERT; ligand;
KW specific; inhibitors; probes; assay; cell sorting; ss.
OS Synthetic.
PN WO9303172-A.
PD 18-FEB-1993.
PF 31-JAN-1992; U00801.
PR 01-AUG-1991; US-739055.
PA (UYRE-) UNIV RES CORP.
PI Gold L, Pribnow D, Smith JD, Tuerk C;
DR WPI; 93-076529/09.
PT Systematic polypeptide evolution by reverse translation - used
PT for prodn. of polypeptide ligand specific for desired target
PT molecule
PS Example 1; Page 84; 98pp; English.
CC SPERT is used to select novel polypeptides that bind the antibody
CC of the epitope commonly recognised by the antisera from autoimmune
CC mice which are the f1 progeny of a cross of NZB and NZW parents
CC (Portanova et al., J. Immunol. 144, 4633, 1990). The known epitope
CC consists of ca. 10 amino acids at the N-terminus of the histone H2B
CC protein. To make mRNA encoding candidate polypeptides a 5' fixed
CC sequence composed of a T7 promoter sequence and a ribosome binding
CC site which is recognised by both prokaryotic and eukaryotic ribosomes,
CC terminating in a restriction endonuclease site is synthesised and cloned
CC using a number of oligonucleotides (example shown). A 3' fixed sequence
CC is placed into a restriction site to provide an mRNA encoding the C-
CC terminal trailer sequence of ca. 100 nucleotides lacking stop codons.
CC In addition, a 3' primer annealing site is provided so that cDNA
CC synthesis can be accomplished on the mRNA recovered from partitioned
CC ribosome complexes. See also Q36845-63.
SQ Sequence 390 BP; 125 A; 126 C; 133 G; 6 T;

Query Match 4.0%; Score 50.8; DB 1; Length 390;
Best Local Similarity 46.2%; Pred. No. 0.014;
Matches 169; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```
Qy    447 CGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGC 506
Db    366 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 307

Qy    507 CCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGG 566
Db    306 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 247

Qy    567 GGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAA 626
Db    246 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 187

Qy    627 CGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGA 686
Db    186 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 127

Qy    687 CAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGC 746
Db    126 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGT 67

Qy    747 CTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCC 806
Db     66 CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCCA 7

Qy    807 AGGCGC 812
Db      6 TGGCCC 1
```

RESULT 13
T68715
ID T68715 standard; DNA; 15872 BP.
AC T68715;
DT 01-SEP-1997 (first entry)
DE Streptomyces venezuelae polyketide synthase vep ORF1.
KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW polyhydroxybutyrate; biodegradable polymer; vep gene;
KW metabolic engineering; ss.
OS Streptomyces venezuelae.
FH Key Location/Qualifiers
FT cds 20. .13912
FT /*tag= a
FT cds 14056. .14136
FT /*tag= b
FT cds 14148. .15827
FT /*tag= c
PN WO9722711-A1.
PD 26-JUN-1997.

PF 18-DEC-1996; U20119.
PR 19-DEC-1995; US-008847.
PA (MINU) UNIV MINNESOTA.
PI Sherman DH, Williams MD, Xue Y;
DR WPI; 97-341701/31.
DR P-PSDB; W19629-30 AND W00918.
PT Expression cassettes for production of polyhydroxyalkanoate(s) -
PT provide wide range of biodegradable polymers for medical or
PT industrial use
PS Claim 54; Fig 23; 91pp; English.
CC Streptomyces venezuelae vep ORF1 (T68715) comprises the polyketide
CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC also W19629-30 and W00918). It contains 5 PKS modules, with a 5'
CC loading module and a 3' end domain. Each of the sequenced modules
CC includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC reductase and an acyl carrier protein domain. The gene cluster was
CC cloned using a heterologous hybridisation strategy from a genomic
CC DNA library. A novel expression cassette encoding the first module
CC from the vep gene cluster and module 7 from the Streptomyces tylP
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.
SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T;

Query Match 4.0%; Score 50.4; DB 1; Length 15872;
Best Local Similarity 47.7%; Pred. No. 0.023;
Matches 187; Conservative 0; Mismatches 196; Indels 9; Gaps 1;

```
Qy    438 CGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCG 497
Db   1540 CCTCCTCGACGGCCTCGGCGCCCTCGCCGCCGGGACGCCCGCGCCCGGCGTGGTCACCGG 1599

Qy    498 ACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCT 557
Db   1600 CACCCCCGCCCCCGGGCGCCTCGCCGTCCTGTTCAGCGGCCAGGGTGCCCAACGTACGGG 1659

Qy    558 GATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGC 617
Db   1660 CATGGGCATGGAGTTGTACGCCGCCCACCCCGCCTTCGCGACGGCCTTCGACGCCGTCGC 1719

Qy    618 CGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTT-------- 670
Db   1720 CGCCGAACTGGACCCCCTCCTCGACCGGCCCCTCGCCGAACTCGTCGCGGCGGGCGACAC 1779

Qy    670 -CTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCA 728
Db   1780 CCTCGACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTGGAGGTCGCCCTCCACCG 1839

Qy    729 CACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGAT 788
Db   1840 CCTCGTCGAGTCCTGGGGCGTCACGCCCGACCTGCTCGCCGGCCACTCCGTCGGCGAGAT 1899

Qy    789 TGGGGCCGTCCTGGTGCCAGGCGCCAAGGCAC 820
Db   1900 CAGCGCCGCCCACGTCGCCGGGGTCCTGTCGC 1931
```

RESULT 14
Q03065
ID Q03065 standard; DNA; 1269 BP.
AC Q03065;
DT 10-MAR-1993 (revised)
DT 06-JUN-1989 (first entry)
DE Sequence encoding Eimeria polypeptide.
KW Eimeria; avian immunisation; coccidosis; ss.
OS Eimeria acervulina.
FH Key Location/Qualifiers
FT cds 1. .1270
FT /*tag= a
PN EP-349071-A.
PD 03-JAN-1990.
PF 23-JUN-1989; 201659.
PR 27-JUN-1988; NL-001627.
PA (ALKU) Akzo NV.
PI Vermeulen A, Dijkema R, Kok JJ;
DR WPI; 90-009080/02.
DR P-PSDB; R04138.
PT Eimeria nucleic acid sequences and polypeptide(s) - used in prepn. of
PT prods. for immunisation of birds, esp. poultry, against coccidosis.
PS Disclosure; Fig 2; 15pp; English.
CC When this sequence or fragment thereof is inserted into a plasmid vector
CC and an E.coli cell transformed with this, resulting polypeptide product
CC can be used to produce a vaccine for immunising birds against coccidosis.
CC In addition to the polypeptides (or fragments) microorganisms expressing
CC these, anti-idiotype antibodies or antigen-binding fragments can be used.
SQ Sequence 1269 BP; 290 A; 348 C; 355 G; 276 T;

Query Match 3.9%; Score 49; DB 1; Length 1269;
Best Local Similarity 43.9%; Pred. No. 0.036;
Matches 257; Conservative 0; Mismatches 325; Indels 3; Gaps 1;

```
Qy    120 ATTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGT 179
Db    147 AGTCATGGTGCCCATTGCCCCGAAGTTTGTTCCAAGGCTCAGAAAGATGGCATTCCGTGT 206

Qy    180 GCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGC 239
Db    207 CATCGTCGAGTCCGGTGCTGGCGCGAATGCTGGCTTCAGTGACGAAGAGTACAGAAGAGC 266

Qy    240 AGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAA 299
Db    267 TGGAGCAGAGATTGCGTCCAACGCCGATGCAGTCATCAACGGAGCTGAGGTGCTGCTCCG 326

Qy    300 GGTCAAAGAACCGATAGCGGCCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCAC 359
Db    327 CGTGTCAGCCCCAACACCGGAAATGGTCTCCCGCATGCCCAGAGACAAGGTGCTGATCAG 386

Qy    360 GTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCAC 419
Db    387 CTACCTCTTCCCCAGCGT AACACGCAAGCATTAGACATGCTTGCACGTCAAGGAGTCAC 446

Qy    420 GTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGAT 479
Db    447 AGCCCTTGCTGTGGACGAAGTGCCACGTGTCACCAGAGCACAA---AAGCTAGACGTTAA 503

Qy    480 GAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCA 539
Db    504 GTCTGCGATGCAAGGCCTCCAGGGCTATCGCGCAGTCATTGAAGCATTCAACGCACTCCC 563

Qy    540 AGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGT 599
Db    564 AAAGCTCAGCAAGGCGTCCATCAGCGCTGCTGGCCGTGTTGAGGCTGCCAAGGTTTTCGT 623

Qy    600 GATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGAC 659
Db    624 TATCGGTGCCGGTGTTGCCGGTCTCCAGGCTATTTCAACTGCCCATGGTTTGGGTGCACA 683

Qy    660 CGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGA 704
Db    684 AGTTTTCGGTCATGATGTGCGCTCCGCAACACGCGAGGAGGTCGA 728
```

RESULT 15
X34205
ID X34205 standard; DNA; 1413 BP.
AC X34205;
DT 06-JUL-1999 (first entry)
DE Mycobacterium species nucleic acid sequence 41S.
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
OS Mycobacterium sp.
PN WO9909186-A2.
PD 25-FEB-1999.
PF 14-AUG-1998; F01813.

PR 11-SEP-1997; FR-011325.
PR 14-AUG-1997; FR-010404.
PA (INSP) INST PASTEUR.
PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y,
PI Guigueno A;
DR WPI; 99-181045/15.
DR P-PSDB; Y04952.
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
PS Claim 22; Fig 41S; 309pp; French.
CC Sequences X34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods
CC for detecting and identifying mycobacteria, especially belonging to
CC the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
SQ Sequence 1413 BP; 187 A; 459 C; 540 G; 227 T;

Query Match 3.8%; Score 47.6; DB 1; Length 1413;
Best Local Similarity 43.0%; Pred. No. 0.072;
Matches 233; Conservative 0; Mismatches 309; Indels 0; Gaps 0;

```
Qy   120 ATTCCGGGTGGCCATCACCCCGGCCGGCGTCGCGGAACT.ACCCGTCGTGGCCATGAGGT 179
Db    99 ATTGCGGGCGGTCGACGCCAACGCCCTCGGGGCCGCGTTGTGGCGCGGCGTCGAGTTGGT 158

Qy   180 GCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGC 239
Db   159 CGTCGCGTCGATGGGTGGCGTGGAGGTGCCGGGAACTATCGTCTCGGTGCTGCGGGCCGC 218

Qy   240 AGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAA 299
Db   219 CGCCGGAGCCGTCGACCAGTGCGCGCACGAGGGGTTGGCCGGTGCGGTCACCGCCGCCGG 278

Qy   300 GGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCAC 359
Db   279 TGACGCGGCGGTCATCGCGCTGGAAAAGACCCCCGAACAGCTTGACGTGCTCGCCGATGC 338

Qy   360 GTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCAC 419
Db   339 GGGCGCGGTGGACGCCGGCGGACGGGGCCTGCTGGTTCTGCTGGACGCGTTGCGCTCCAC 398

Qy   420 GTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGAT 479
Db   399 CATCTGCGGGCAGGCACCTGCCCGGGCGGTCTACGAACCCTCGCCGCGCGCGTTGCCGAC 458

Qy   480 GAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCA 539
Db   459 CGACACGGCTACCCAACGCCCCGCCCCGCAATTCGAGGTGATGTATCTGTTGGCGGTATG 518

Qy   540 AGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGT 599
Db   519 TGATGCTGCAGCGGCGGACCAGTTGCGGGATCGACTCAAGGAATTGGGTGAGTCGGTGGC 578

Qy   600 GATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGAC 659
Db   579 CATCGCCGCTGCTCCGCCCGACAGCTACTCCGTACACGTCCACACCGACGACGCCGGTGC 638

Qy   660 CG 661
Db   639 CG 640
```

Search completed: June 22, 2000, 15:04:40
Job time: 17606 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 10:08:42 ; Search time 341.15 Seconds
(without alignments)
480.085 Million cell updates/sec

Title: US-09-362-485-1
Perfect score: 1260
Sequence: 1 GAATTCCCATCAGCAATCTT.....AAGCGATGATGTCGGCCGCG 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	47.6	3.8	1620	3	US-08-461-775-10	Sequence 10, Appl
2	47.6	3.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
3	47.6	3.8	28958	1	US-08-456-837-6	Sequence 6, Appli
4	47.6	3.8	28958	1	US-08-457-342-6	Sequence 6, Appli
5	47.6	3.8	28958	1	US-08-457-646A-6	Sequence 6, Appli
6	47.6	3.8	28958	2	US-08-458-076A-6	Sequence 6, Appli
7	47.6	3.8	28958	2	US-08-764-233A-4	Sequence 4, Appli
8	47.6	3.8	28958	2	US-08-457-335A-6	Sequence 6, Appli
9	47.6	3.8	28958	2	US-08-729-214-6	Sequence 6, Appli
10	47.6	3.8	49377	2	US-08-764-233A-1	Sequence 1, Appli
11	46	3.7	2582	4	US-08-816-105A-2	Sequence 2, Appli
12	45.4	3.6	833	3	US-08-403-852D-3	Sequence 3, Appli
13	45.4	3.6	5392	3	US-08-403-852D-1	Sequence 1, Appli
14	42.4	3.4	1833	3	US-08-403-852D-6	Sequence 6, Appli
15	42.4	3.4	2185	1	US-08-173-508-3	Sequence 3, Appli
16	42.4	3.4	2185	3	US-08-265-310-3	Sequence 3, Appli
17	41.2	3.3	2414	7	5248599-1	Patent No. 5248599
18	41.2	3.3	2668	3	US-08-461-775-11	Sequence 11, Appl
19	41	3.3	20235	2	US-07-642-734C-3	Sequence 3, Appli
20	40.8	3.2	44377	3	US-08-804-227C-7	Sequence 7, Appli
21	40.8	3.2	44377	4	US-08-804-198-1	Sequence 1, Appli
c 22	40.6	3.2	4257	3	US-08-690-473-1	Sequence 1, Appli
23	40.6	3.2	12001	2	US-08-458-568A-11	Sequence 11, Appl
24	40.2	3.2	2064	1	US-08-343-428-1	Sequence 1, Appli
25	40.2	3.2	11219	2	US-07-642-734C-1	Sequence 1, Appli
26	39.6	3.1	474	3	US-08-403-852D-14	Sequence 14, Appl
27	39.4	3.1	459	4	US-08-387-942C-35	Sequence 35, Appl
c 28	39.4	3.1	1998	7	5212296-8	Patent No. 5212296
29	39.4	3.1	12588	4	US-08-387-942C-1	Sequence 1, Appli
c 30	39.2	3.1	420	1	US-08-470-179-148	Sequence 148, App
c 31	39.2	3.1	2353	6	PCT-US92-06840-1	Sequence 1, Appli
32	39	3.1	43280	3	US-08-804-227C-1	Sequence 1, Appli
33	38.8	3.1	1524	7	5512669-1	Patent No. 5512669
34	38.8	3.1	1525	7	5229279-1	Patent No. 5229279
35	38.8	3.1	8051	4	US-08-576-626A-2	Sequence 2, Appli
c 36	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appli
37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appli
38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appli
39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 4, Appli
c 40	38.4	3.0	1215	4	US-08-947-726A-1	Sequence 1, Appli
41	38.2	3.0	2109	4	US-08-555-568B-20	Sequence 20, Appl
c 42	38	3.0	30001	1	US-08-125-468-1	Sequence 1, Appli
c 43	38	3.0	30001	3	US-08-474-933-1	Sequence 1, Appli
44	37.8	3.0	2588	3	US-08-796-414B-6	Sequence 6, Appli
c 45	37.8	3.0	15664	1	US-08-402-282-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT   1
US-08-461-775-10
; Sequence 10, Application US/08461775
; Patent No. 5858773
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,775
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1620 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1620
US-08-461-775-10
```

```
  Query Match             3.8%;  Score 47.6;  DB 3;  Length 1620;
  Best Local Similarity   45.7%;  Pred. No. 0.0043;
  Matches  321;  Conservative    0;  Mismatches  369;  Indels   12;  Gaps    4;

Qy     135 CACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGG 194
           |||||   |||| | | | |   |  ||      | |        ||||  |||||| ||
Db     879 CACCCTCACCGGTGCCACCGTCATCGCCGAGGAGGTCGGCCTCAAGCTCGACCAGGCCGG 938

Qy     195 TGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGT 254
           |   | | |       |   | | | ||      |||    ||  | | ||    | ||
Db     939 TCTGGACGTGCTGGGCACCGCCCGCCGCGTCACCGTCACCAAGGACGACACGACCATCGT 998

Qy     255 CGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGAT 314
           | | || || | | |    || || |   |       |   | || |||| |    |||
Db     999 GGACGGCGGCGGCAACGCCGAGGACGTCCAGGGCCGCGTCGCCCAGATCAAGGCCGAGAT 1058

Qy     315 AGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGC 374
           | | ||   ||    |  |  | || | || ||  |        ||    |  ||||
Db    1059 CGAGTCGACCGACTCGGACTGGGACCGCGAGAAGCTCCAGGAGCGCCTCGCCAAGCTGGC 1118

Qy     375 C-GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGCCACCACGTCAATTGCCTACG 433
           | ||| |   ||| ||  |||   ||  |   |   | | | | ||    |
Db    1119 CGGCGGCGTCTGCGTGATCCGCGTCGGCGCGGCCACCGAGGTCGAGCTGAAGGAGCGCAA 1178

Qy     434 AGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAA--GTCGC 491
             ||||||  ||   |        | | ||||  ||      ||| ||| | |  ||| |
Db    1179 GCACCGTCTGGAGGACGCCATCTCCGCGACCCGCGCCGCGGTCGAGGAGGGCATCGTCTC 1238

Qy     492 CGGTCGACTCGCCGCCCAGGTTGGCGCTTACCA---CCTGATGCGAACCCAAGGGGGCCG 548
           |||| |   | |||| | |||   |||   | |   ||||      | ||  ||  ||
Db    1239 CGGTGGTGGCTCCGCGCTGGTCCACGCCGTCAAGGTCCTGGACGACAACCTCGGCCGCAC 1298

Qy     549 CGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG--- 606
           ||| |    |     ||| ||  | | ||||   |  |||| ||||| |  | |  |
Db    1299 CGGCGACGAGGCCACCGGTGTCGCGGTCGTCCGCCGCGCCGCCGTCGAGCCGCTGCGCTG 1358

Qy     606 ---CGCCGGCACCGCCGGCTACAACGCAGCCCGCATCG CAACGGCATGGGCGCGACCGT 662
              |||||  | |||||||  | | |    |  |||| ,|| |    |||  | |  ||
Db    1359 GATCGCCGAGAACGCCGGCCTCGAGGGCTACGTCATCACCACCAAGGTGGCGGAGCTCGA 1418

Qy     663 TACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCG 722
           | ||  |  | | |||||   | ||       | |  ||||   |       ||||
Db    1419 CAAGGGCCAGGGCTTCAACGCGGCCACCGGCGAGTACGGCGACCTGGTCAAGGCCGGCGT 1478

Qy     723 GATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCT 782
           ||| || |   | |     | || | ||  || |||   || | |      |||  |||
Db    1479 CATCGACCCGGTCAAGGTCACCGCGTCCGCCCTGGAGAACGCGGCCTCCATCGCCTCCCT 1538

Qy     783 GGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAA 824
           | |  |   | |||      ||   | ||  ||| |  || |
Db    1539 GCTCCTGACGACCGAGACCCTGGTCGTCGAGAAGCCGGCCGA 1580


RESULT   2
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,261B
;      FILING DATE:  08-JUN-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6


  Query Match             3.8%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy     169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
           | ||| || |  |||| ||  || | ||  |   |      ||| | ||||  |||  |
Db   25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy     229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
           ||| ||       | ||| |   | | |  || |  |||| || |     |||    ||
Db   25366 TTCGAGCACCGCGGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy     289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
           |   |  | ||| |||      |        | |   ||| ||     ||  |
Db   25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy     349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
           |   |  ||    |||| | |  |   |  || |                ||  | | |
Db   25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy     409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
             || | || || |  |    |  ||       || |  |   |||  |||| ||  | |
Db   25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy     469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
           |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | || |  |
Db   25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy     527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
             |   |||   |     | ||||  | |          ||  | || |||| |   | ||
Db   25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy     587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
           |   || |||  ||      ||||| |  || |||| |||     | |   ||||||  |
Db   25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy     646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
```

```
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   3
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6

  Query Match             3.8%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228

Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCAGCCGCGCCCAC 25365

Qy    229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288

Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348

Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408

Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468

Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526

Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586

Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645

Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705

Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   4
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6


  Query Match             3.8%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy     169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
           | ||| || |  |||| ||  || | ||  |   |      ||| | ||||  |||  |
Db   25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy     229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
           ||| ||      | ||| |   | | |  || |  |||| || |     |||     ||
Db   25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy     289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
           |  |   | ||| |||     |        | |    ||| ||      ||  |
Db   25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy     349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
           |   |  ||    |||| | |  |   |  || |                ||  | | |
Db   25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy     409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
             || | || || |  |   |  ||      || |  |   |||  |||| ||  | |
Db   25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy     469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
           |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | || |  |
Db   25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy     527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
             |  |||  |     | ||||  | |          ||  | || |||| |   | ||
Db   25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy     587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
           |   || |||  ||      ||||| |  || |||| |||     | |   ||||||  |
Db   25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy     646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
           | | | | |||   || |      || ||   |      |  ||        |  | |  |
Db   25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy     706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
             | | ||  |  |      |  ||   |    |    | |||  |||||||  |   ||
Db   25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGACTCGGCGCCTCC 25905
```

```
Qy     766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
           | |   |   || |||| |      ||| || |  | |  |   ||||
Db   25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   5
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6


  Query Match             3.8%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy     169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
           | ||| || |  |||| ||  || | ||  |   |      ||| | ||||  |||  |
Db   25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy     229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
           ||| ||      | ||| |   | | |  || |  |||| || |     |||     ||
Db   25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425
```

```
Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTA AACGCAGCCCGCATCGCCAAC 645
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT   6
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6


  Query Match             3.8%;  Score 47.6;  DB 2;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
Db  25486 AAGCTCGTCTTCGTCTT CCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   7
US-08-764-233A-4
```

```
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1
US-08-764-233A-4


  Query Match             3.8%;  Score 47.6;  DB 2;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy     169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
           | ||| || |  |||| ||  || | ||  |   |      ||| | ||||  ||| |
Db   25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy     229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
           ||| ||       | ||| |   | | |  || |  |||| || |     |||     ||
Db   25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy     289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
           |   |   | ||| |||      |      | |    ||| ||     ||  |
Db   25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy     349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
           |   | ||    |||  | |  |   |  || |                ||  | | |
Db   25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTG GAAGGGATGGCCCTCTCGCTG 25545

Qy     409 TCCGGCACCACGTCAAT GCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
           || | || || |  |   | ||      || |  |  |||  |||| ||  | |
Db   25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy     469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
           |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | || |  |
Db   25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy     527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
           |   |||   |     | ||||  | |           ||  | || |||| |   | ||
Db   25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy     587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
           |   || |||  ||       ||||| |  || |||| |||     | |  ||||||  |
Db   25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy     646 GGCATGGGCGCGACCGTCACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
           | | | |||   || |      || ||   |   ||       | |||  | | |
Db   25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy     706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
           |  | ||  |  |    | ||   |   |    | |||  |||||||| |   ||
Db   25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy     766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
           | |   |  || |||| |      ||| || |  | |  |   ||||
Db   25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   8
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,335A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6
```

```
  Query Match             3.8%;  Score 47.6;  DB 2;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
          | ||| || |  |||| ||  || | ||  |    |       ||| | ||||  |||  |
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCCCTGGCCACCACCCGCGCCCAC 25365

Qy    229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGAC CAGGTGTGGGCCGACGCTG  288
          ||| ||       | ||| |   | | |  || |  |||| || |     |||     ||
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
          |   |   | ||| |||     |        | |    ||| ||     ||  |
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
          |   |  ||    |||| | |  |   |  || |                 ||  | | |
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
            || | ||  || |  |   |  ||       || |  |   |||  |||| ||  | |
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGC \TGCGAGCGCGCGCTCGCTCCT 25605

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
          |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | ||  |  |
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
            |   |||   |      | ||||  | |           ||  | || |||| |    | ||
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
          |   || |||  ||      ||||| |  || |||| |||     | |   ||||||  |
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
          | | | | |||   || |      || ||     |  ||        |  | | |  |
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
            | | ||  |  |     |  ||   |  |    | |||  |||||||  |   ||
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
          | |   |  || |||| |      ||| || |  | |  |   ||||
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT   9
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6
```

```
  Query Match             3.8%;  Score 47.6;  DB 2;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0098;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228
          | ||| || |  |||| ||  || | ||  |    |       ||| | ||||  |||  |
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288
          ||| ||       | ||| |   | | |  || |  |||| || |     |||     ||
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348
          |   |   | ||| |||     |        | |    ||| ||     ||  |
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
          |   |  ||    |||| | |  |   |  || |                 ||  | | |
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
            || | ||  || |  |   |  ||       || |  |   |||  |||| ||  | |
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
          |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | ||  |  |
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
            |   |||   |      | ||||  | |           ||  | || |||| |    | ||
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
          |   || |||  ||      ||||| |  || |||| |||     | |   ||||||  |
```

```
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705

Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT  10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologous
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs t
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS ge
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  46851..47891
;      OTHER INFORMATION:  /product= "SorM"
;      OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;      OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;      OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;      OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1
```

```
  Query Match             3.8%;  Score 47.6;  DB 2;  Length 49377;
  Best Local Similarity   44.1%;  Pred. No. 0.011;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    169 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 228

Db  41609 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 41668

Qy    229 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 288

Db  41669 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 41728

Qy    289 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 348

Db  41729 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 41788
```

```
Qy    349 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 408
Db  41789 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 41848

Qy    409 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 468
Db  41849 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 41908

Qy    469 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 526
Db  41909 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 41968

Qy    527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
Db  41969 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 42028

Qy    587 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
Db  42029 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 42088

Qy    646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
Db  42089 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 42148

Qy    706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
Db  42149 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 42208

Qy    766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 813
Db  42209 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 42256
```

```
RESULT  11
US-08-816-105A-2
; Sequence 2, Application US/08816105A
; Patent No. 5989882
;  GENERAL INFORMATION:
;    APPLICANT:  Crueger, Anneliese; Dellweg, Hans-Georg; Lenz,
;    APPLICANT:  Jurgen; Schroder, Werner; Pape, Hermann;
;    APPLICANT:  Goeke, Klaus; Schaper, Beate; Hemker, Michael;
;    APPLICANT:  Piepersberg, Wolfgang; Distler, Jurgen;
;    APPLICANT:  Stratmann, Ansgar
;    TITLE OF INVENTION:  PROCESSES FOR PREPARING ACARVIOSYL
;    TITLE OF INVENTION:  TRANSFERASE AND FOR USING IT IN THE
;    TITLE OF INVENTION:  CONVERSION OF ACARBOSE HOMOLOGUES
;    TITLE OF INVENTION:  INTO ACARBOSE, FOR THE PREPARATION
;    TITLE OF INVENTION:  OF ACARBOSE HOMOLOGUES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh 6500
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect 3.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,105A
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19625269.5
;      FILING DATE:  25-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19611252.4
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9814-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2582 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-816-105A-2
```

```
  Query Match             3.7%;  Score 46;  DB 4;  Length 2582;
  Best Local Similarity   47.3%;  Pred. No. 0.012;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy    374 CCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACG 433
Db   1576 CCGCGCTGCCGGCCGGCACCTATCGCGACGTGCTCGGCGGCACCTTCGGCGGCTCCGACC 1635

Qy    434 AGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCG 493
Db   1636 TGACCGTCGAGGACGGCGACGGCACCGACCGGTCGACCGTCGCGCCGGTGCTGGGTGCCG 1695

Qy    494 GTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTG 553
Db   1696 GGCAGGTCGCCGTCTGGTCGTACCGGGCGCCGGTGGACACCGAGCCCCGGATCGGCGGGG 1755

Qy    554 TGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCA 613
Db   1756 TCGGGCCGGTCGTGACCCGGGCCGGCGCCACCGTCACCGTCGAGGGCACCGGCTTCGGCT 1815

Qy    614 CCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGG 667
Db   1816 CCGGCGGAACCGTCGCGATCGGCGGAGTCCCCGCGACCGTCCAGCAGTGGACGG 1869
```

```
RESULT  12
US-08-403-852D-3
; Sequence 3, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  PCT/FR 93/00923
;       FILING DATE:  25-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 92/11441
;       FILING DATE:  25-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meyers, Kenneth J.
;       REGISTRATION NUMBER:  25,146
;       REFERENCE/DOCKET NUMBER:  03806.0054-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 408-4000
;       TELEFAX:  (202) 408-4400
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  833 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  S.pristinaespiralis
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..833
US-08-403-852D-3


  Query Match             3.6%;  Score 45.4;  DB 3;  Length 833;
  Best Local Similarity   48.0%;  Pred. No. 0.012;
  Matches  130;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     564 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 623
           || ||  | |  || |||||    |||||||||       |      | || |  || | |
Db     327 CGAGGGCCGCACCGGCGAACTCATCGACGTCGTCCGCGGCCTGTGGGACAGCTTCGACGA 386

Qy     624 CAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACAT 683
           | ||||   |  |  || |  || |   |||  |  |     | | |  | | || ||
Db     387 CGACGCCTTCGTCCACGACCGCGCCGACGGCCTGTACTGGCGGCTGCCCGCCGTCCACCA 446

Qy     684 CGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATC 743
           | | |  | |       ||  |||  |   |||||  || ||     |  | |  |
Db     447 ACTCGACCACCAGGGCAGGCACTTCGACGTGGCCGGCCCCCTCAACGTCGCCCGCCCGCC 506

Qy     744 GGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGT 803
           |     | | ||  |  |||||||   |  ||| ||| |     |  |||| ||| ||
Db     507 GCAGGGCCACCCCGTCGTCGCCGTCACCGGCCCCGCCCTCGCCGCGGCCGCCGACCTCGT 566

Qy     804 GCCAGGCGCCAAGGCACCCAAATTAGTCTCG 834
           |    || | ||||  || |    | ||||
Db     567 CCTGCTCGACGAGGCGGCCGACGCCGCCTCG 597


RESULT  13
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristina spiralis
US-08-403-852D-1


  Query Match             3.6%;  Score 45.4;  DB 3;  Length 5392;
  Best Local Similarity   48.0%;  Pred. No. 0.021;
  Matches  130;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     564 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 623
           || ||  | |  || |||||    |||||||||       |      | || |  || | |
Db    3018 CGAGGGCCGCACCGGCGAACTCATCGACGTCGTCCGCGGCCTGTGGGACAGCTTCGACGA 3077

Qy     624 CAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACAT 683
           | ||||   |  |  || |  || |   |||  |  |     | | |  | | || ||
Db    3078 CGACGCCTTCGTCCACGACCGCGCCGACGGCCTGTACTGGCGGCTGCCCGCCGTCCACCA 3137

Qy     684 CGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATC 743
           | | |  | |       ||  |||  |   |||||  || ||     |  | |  |
Db    3138 ACTCGACCACCAGGGCAGGCACTTCGACGTGGCCGGCCCCCTCAACGTCGCCCGCCCGCC 3197

Qy     744 GGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGT 803
           |     | | ||  |  |||||||   |  ||| ||| |     |  |||| ||| ||
Db    3198 GCAGGGCCACCCCGTCGTCGCCGTCACCGGCCCCGCCCTCGCCGCGGCCGCCGACCTCGT 3257

Qy     804 GCCAGGCGCCAAGGCACCCAAATTAGTCTCG 834
           |    || | ||||  || |    | ||||
Db    3258 CCTGCTCGACGAGGCGGCCGACGCCGCCTCG 3288


RESULT  14
US-08-403-852D-6
; Sequence 6, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
```

; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,852D
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1689
US-08-403-852D-6

Query Match 3.4%; Score 42.4; DB 3; Length 1833;
Best Local Similarity 50.5%; Pred. No. 0.083;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy    554 TGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCA 613
          || || | ||   ||||| |  || ||    | ||| |   | ||   ||| |   | |
Db    413 TGATGCTCGGACTGGTGCTCTTCGGCGCCGGGTCCGCCTGGGCGGCCTTCGCCCAGGACT 472

Qy    614 CCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAG 673
          |||||     || ||| |||||   || ||  ||| |||||| | |  |   ||
Db    473 CCGCCCAACTCATCGCCGCCCGGGCCGGCATGGGCGTGGGCGGGGCGCTGCTGGCGACCA 532

Qy    674 ACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTC 733
          || |  |  || ||   | | |    |||||| |||     ||  ||||     |   |
Db    533 CCACCCTCGCCGTCATCATGCAGGTCTTCGACGACGACGAACGCCCCCGGGCGATCGGCC 592

Qy    734 GCTACTCATCGGCCTACGAGCTCG 757
```

```
          |    | |||||  |   || |
Db    593 TGTGGGGAGCGGCCAGCTCACTGG 616
```

RESULT 15
US-08-173-508-3
; Sequence 3, Application US/08173508
; Patent No. 5616485
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
; TITLE OF INVENTION: S' (EPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,508
; FILING DATE: 23-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/125/CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2185 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 531..2069
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 531..902
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 903..2069
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 531..533
; OTHER INFORMATION: /note= "Met at position -124
; OTHER INFORMATION: represents fMet"
US-08-173-508-3

Query Match 3.4%; Score 42.4; DB 1; Length 2185;
Best Local Similarity 47.4%; Pred. No. 0.087;
Matches 127; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy    546 CCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG 605
          | |||  |   |||  |||| ||  ||||| ||  |   ||||| | ||  |||   |
Db   1250 CGGCGTCGGCATGACCGGCGTGGCACCCGGGGTGAAGGTGGCCGGCATCAAGGTCTCCAA 1309

Qy    606 CGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTAC 665
```

```
Db   1310 CCCCGACGGCTTCTTCTACACCGAGGCCGTGGTCTGCGGCTTCATGTGGGCGGCCGAGCA 1369

Qy    666 GGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGAT 725

Db   1370 CGGCGTCGACGTGACCAACAACAGCTATTACACCGACCCGTGGTACTTCAACTGCAAGGA 1429

Qy    726 CCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGT 785

Db   1430 CGACCCCGACCAGAAGGCGCTCGTCGAGGCCGTCTCGCGGCCTCCCGGTACGCGGAGAA 1489

Qy    786 GATTGGGGCCGTCCTGGTGCCAGGCGCC 813

Db   1490 GAAGGGCGCGGTCAACGTCGCCGCGGCC 1517
```

Search completed: June 22, 2000, 14:58:42
Job time: 17400 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 09:14:46 ; Search time 5541.94 Seconds
(without alignments)
921.530 Million cell updates/sec

Title: US-09-362-485-1
Perfect score: 1260
Sequence: 1 GAATTCCCATCAGCAATCTT.....AAGCGATGATGTCGGCCGCG 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

	No.	Score	Match	Length	DB	ID	Description
c	1	57.4	4.6	925	82	CNS0091P	AL053013 Drosophil
	2	55.2	4.4	925	82	CNS0091P	AL053013 Drosophil
	3	51.8	4.1	535	74	AV390505	AV390505 AV390505
c	4	50.2	4.0	844	82	CNS0052P	AL056652 Drosophil
c	5	49	3.9	910	82	CNS006ON	AL065629 Drosophil
c	6	46.6	3.7	846	82	CNS010RJ	AL099337 Drosophil
c	7	46.6	3.7	1101	83	CNS0181E	AL108764 Drosophil
	8	46.4	3.7	935	82	CNS006XK	AL066051 Drosophil
c	9	46.2	3.7	645	82	CNS012I3	AL101589 Drosophil
c	10	45.8	3.6	935	82	CNS006XK	AL066051 Drosophil
	11	45.2	3.6	744	83	CNS017ZK	AL108698 Drosophil
	12	45.2	3.6	1101	83	CNS017SY	AL108460 Drosophil
	13	45	3.6	315	51	AI712257	AI712257 605069D09
	14	45	3.6	846	82	CNS010RJ	AL099337 Drosophil
	15	44.8	3.6	932	82	CNS0072Q	AL066742 Drosophil
c	16	44	3.5	843	82	CNS00CS1	AL059666 Drosophil
	17	43.8	3.5	427	51	AI736678	AI736678 sb32b04.y
c	18	43.6	3.5	1101	83	CNS016UE	AL107216 Drosophil
c	19	43.2	3.4	439	91	AQ864063	AQ864063 nbeb0022M
c	20	43.2	3.4	798	82	CNS00AJA	AL055851 Drosophil
c	21	43	3.4	413	74	AV396947	AV396947 AV396947
	22	41.8	3.3	414	80	AW286084	AW286084 LG1_261_B
	23	41.8	3.3	465	80	AW287007	AW287007 LG1_264_C
	24	41.8	3.3	530	74	AV396353	AV396353 AV396353
	25	41.8	3.3	839	82	CNS004NB	AL054280 Drosophil
c	26	41.8	3.3	932	82	CNS0072Q	AL066742 Drosophil
	27	41.6	3.3	506	74	AW180713	AW180713 MgA0863f
c	28	41.6	3.3	870	82	CNS006EZ	AL064271 Drosophil
c	29	41.4	3.3	1101	82	CNS00LXJ	AL078875 Drosophil
	30	41.4	3.3	1201	83	CNS014BJ	AL103945 Drosophil
	31	41.2	3.3	639	60	AI812194	AI812194 605086F10
	32	41.2	3.3	774	105	AQ327321	AQ327321 nbxb0040D
	33	41.2	3.3	910	82	CNS006ON	AL065629 Drosophil
	34	41	3.3	605	46	AI461529	AI461529 486018G05
c	35	41	3.3	1101	83	CNS017SY	AL108460 Drosophil
	36	40.8	3.2	384	64	AW054773	AW054773 ws59d08.x
c	37	40.8	3.2	385	21	D48746	D48746 RICS15137A
c	38	40.8	3.2	437	74	AV395790	AV395790 AV395790
c	39	40.4	3.2	574	91	AQ848186	AQ848186 LMAJFV1_1
c	40	40.2	3.2	916	83	CNS015SQ	AL105860 Drosophil
	41	40	3.2	374	50	AU068004	AU068004 AU068004
c	42	40	3.2	868	85	AQ688189	AQ688189 nbxb0077A
	43	39.8	3.2	525	61	AU075646	AU075646 AU075646
c	44	39.8	3.2	550	59	AI770338	AI770338 606061B10
	45	39.8	3.2	1101	83	CNS017P2	AL108320 Drosophil

ALIGNMENTS

RESULT 1
CNS0091P/c
LOCUS CNS0091P 925 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 4.6%; Score 57.4; DB 82; Length 925;
Best Local Similarity 11.9%; Pred. No. 0.05;
Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

```
Qy    463 CCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTAC 522
          :|:|  : ::|:::::   ::   : ::|:::::: :|:::::::: ::  :::::   :
Db    920 SCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTS 861

Qy    523 CACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAA 582
          :||::   :::  ||     :  :|::::   |: :| |:::: :|     || :   |
Db    860 SACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSGGASA 801

Qy    583 CCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCC 642
          ::::::  :: ::  : : :: ::::: ::  ::: ||   : |::::::::::  :  :
Db    800 SHSSSSACBSSSSSCSAS WSASSSSSASSSRSRSGGGAGGS ASSSRSSSSSSSASAGS 741

Qy    643 AACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTC 702
          :::: :  :::|:::::|:  |::   :  : :  :| :  :: :|     :  |   |
Db    740 VVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCCTSWSC 681

Qy    703 GACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGT 762
          : : : :::    :::::::   ::::|: : :  : :  ::: :  :: :   : ::
Db    680 SCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBSMSSG 621

Qy    763 GCCGTCAAACGTGCCGACCTGGTGA TGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCC 822
          | :| : | :| :::: :: |: :   ||||: |  : :: : : ||:|::   :: |:|
Db    620 GGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSSGCMC 561

Qy    823 A 823
          :
Db    560 R 560
```

RESULT 2
CNS0091P
LOCUS CNS0091P 925 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN
```

```
  Query Match             4.4%;  Score 55.2;  DB 82;  Length 925;
  Best Local Similarity   11.8%;  Pred. No. 0.13;
  Matches   41;  Conservative  170;  Mismatches  137;  Indels    0;  Gaps    0;

Qy     446 CCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCG 505
           :|:|:: : :| : |  |:|:  |  :::::: :  ::  : | :::: || : :::::
Db     577 SCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTSBSCSCCCCSSKSVCGTSCSSSSS 636

Qy     506 CCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCG 565
           |:: :: |::::     : :: :  :::  :::  ::: : :|:|::: |  :   |: :
Db     637 CSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTS 696

Qy     566 GGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACA 625
           ::: ::: : :: ::   :::: : :| :: :| :  :: ::::::    :::::::: :
Db     697 SSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSSSYSSS 756

Qy     626 ACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCG 685
           |:|  |||:: :::::   :::  :|: :|:: ::: | ::::::      :  | :|  |:
Db     757 TCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCS 816

Qy     686 ACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGG 745
           : : |:|: :|  : ::  :  : : | |||:::    :  :   :::  : :  :::
Db     817 TYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCSSSSVS 876

Qy     746 CCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGG 793
           ::  ::|:: :: ::| : ::      :::::: :: :| |   |:::
Db     877 SSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVS 924
```

```
RESULT   3
AV390505
LOCUS       AV390505      535 bp    mRNA            EST       21-JAN-2000
DEFINITION  AV390505 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            cDNA clone CM052b05_r, mRNA sequence.
ACCESSION   AV390505
VERSION     AV390505.1  GI:6544721
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
  TITLE     A Large Scale Structural Analysis of cDNAs in a Unicellular Green
            Alga, Chlamydomon s reinhardtii. I. Generation of 3433
            Non-redundant Expressed Sequence Tags
  JOURNAL   DNA Res. 6, 369-373 (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135919.
            Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="CM052b05_r"
                     /clone_l b="Chlamydomonas reinhardtii C9"
                     /dev_stage="photoautotrophic growth"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      107 a    195 c    148 g     85 t
ORIGIN
```

```
  Query Match             4.1%;  Score 51.8;  DB 74;  Length 535;
  Best Local Similarity   51.5%;  Pred. No. 0.57;
  Matches  119;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     586 GCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 645
           || |||| |||| |||| || || ||  |  | |  |||||| |  || || |
Db     221 GCTGACGAGGTGGCCATCGAGACCTGCTCCTTCTCCAAGTACGCAGGCTTCACCGGCGTG 280

Qy     646 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 705
           || |||||  ||||||  |      | | ||||  || |||      ||  |  ||
Db     281 CGCCTGGGCTGGACCGTGGTGCCCAAGGCCCTCAAGTACGCCAACGGCGAGCCGGTGCAC 340

Qy     706 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 765
           || || |    | ||   ||   |||  ||| |      |||| | |  ||| |   ||
Db     341 GCGGACTGGAACCGCGTCATGACCACCTGCTTCAACGGCGCCTCCAACATCGTGCAGGCG 400

Qy     766 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAG 816
           | |   | |||  |||| |   | ||   ||  || |     |||| |
Db     401 GGCGGTCTGGCCTGCCTGCAGCCGGAGGGGCTCAAGGAGATGAACGCCATG 451
```

```
RESULT   4
CNS0052P/c
LOCUS       CNS0052P      844 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR11P16"
                     /note="end : TET3"
BASE COUNT      261 a    112 c     92 g     35 t    344 others
ORIGIN
```

```
  Query Match             4.0%;  Score 50.2;  DB 82;  Length 844;
  Best Local Similarity   17.2%;  Pred. No. 1.2;
  Matches   65;  Conservative  161;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     278 CCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGC 337
           || :: ||  |:| |: | ::|   : |: :::  ::: : ::  |: ::   : :  ::
Db     768 CCMMVCCTTTTHTTTYTCMYWAMMRMYAWMMMMAMMWVTMCVMMMAVAMVMYAMAMAMRM 709

Qy     338 GACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATG 397
           : |:::| : | |:   : |:  :: :: ||   :  : |: ::| :|| :: :  |
Db     708 NNVYCSBSCNBBTTTYYCNYTCVSSSBSSSBTTCWTHTNSNTYTMVTYVTTTHHTYTNTC 649

Qy     398 CGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCG 457
           | :       : :| :: | :  :  : :|:  :: : ::  :  : :: ||: | :
Db     648 CYYCCCCCCYYCSCCBYBCNTSCSYYSYSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGY 589

Qy     458 CACTAC-CCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGC 516
           | |: : |:::::|::::: :  : ::|   :|::::: :: : :::| |:| :  |  :
Db     588 CGCYSBVCSSBSSTBSSVSGSBSSTSSGTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCY 529

Qy     517 GCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGC 576
           ::   :| ::   |::   |::   |::::| :: |::: :  |:::||:|: :|:  |
Db     528 SSCCBYCCBSGCBTBSYBSCSSGBTGBSSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYC 469

Qy     577 GTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGC 636
             ::  ||:::::  ::|  || : |  |::::|| ::: :|::  :  ::: |:: ::
Db     468 CSBSSBCCSSSSSSTSKCCSGGSSSTKCGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSS 409

Qy     637 ATCGCCAACGGCATGGG 653
           :::::  ::::  | |
Db     408 KKYSYSTSSSSSGCGTG 392
```

```
RESULT   5
CNS006ON/c
LOCUS       CNS006ON      910 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065629
VERSION     AL065629.1  GI:4944698
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14J21"
                     /note="end : T7"
BASE COUNT      202 a     63 c    112 g    198 t    335 others
ORIGIN
```

```
  Query Match             3.9%;  Score 49;  DB 82;  Length 910;
  Best Local Similarity   20.5%;  Pred. No. 2.1;
  Matches   47;  Conservative   97;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     414 CACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGC 473
           ::::|:  : |  :::   : :|::: ::|::::| :|::: : |: | :    :  :
Db     713 SRVMASAGSSASSSSSSSASSSASSSASSASMVVSCRSAVSVASAASAAASARSASRASA 654

Qy     474 CCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCG 533
           :: :|  |:::|:  :|:::: |: : |||: :::  :  :|:::   :: :|      :
Db     653 VMAVAVAASSSAVASSGSSSVSCSSSRCGCSASSSSASASSGSSSSSSSSSSCGCGSCAS 594

Qy     534 AACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGT 593
             |::: ::|: || :::| : :: |  :::::::: : |::::  |   :|  |   |
Db     593 CASSSSRSGSCGCGSSVGCVSSSCGSSSSSSSSSSSSGCSSSSCCCCSCSCACCRCCCCC 534

Qy     594 CGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCC 642
           |:  : |  | :::||::| || || ::   :|: :| ||: |  |:||
Db     533 CRSCSCGCCCMSSSCCSSCCCCCCCCASMCCAWAMAHACCCSCCCCCDCC 485
```

```
RESULT   6
CNS010RJ/c
LOCUS       CNS010RJ      846 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099337
VERSION     AL099337.1  GI:5610948
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
```

```
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04N13"
                     /note="end : T7"
BASE COUNT       300 a     47 c     40 g    175 t    284 others
ORIGIN
```

```
  Query Match             3.7%;  Score 46.6;  DB 82;  Length 846;
  Best Local Similarity   13.0%;  Pred. No. 6.2;
  Matches   33;  Conservative  127;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     384 TGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCA 443
           | || :: ::   :::   :|:  ||: ::   : :: :    :::| :  : : :: :
Db     842 TTCTSSSTSSTSCSSSSSTBTSTTTTSTSSTSASSSSSSSTSSSSSTSSTSSSSTSBTBT 783

Qy     444 GACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGC 503
           : ::  :   ::::  :| ::::| :| :|::::   ::: |: :: : :    | :::
Db     782 SBSSTTSTTTSSSSTTSTTSSSSTCSTSSCSSSSSSTTSSSSASASSTSTSSTTGCSBSS 723

Qy     504 CGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGG 563
           ::: :  : |::: :||  : ::::  :::   ::  : ::::::::|:|::    ::
Db     722 SBSTSGTSSTSBBTBTTBTSSSSBBSRSSSSSTSSTTBTSBSSSSSBSTSTSSSTTSTSS 663

Qy     564 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 623
           :::::|:: :::::||      :::: :: ::|::|:  ::::::|: : |:::::::|
Db     662 SSSSBTSSTSSBSSTCTTSTTTSSSSCSSSSSTSSTSSSBSSSSSCBTSSCSSSSSSCCT 603

Qy     624 CAACGCAGCCCGC 636
           |  |:|  ||| |
Db     602 CGCCSCSCCCCCC 590
```

```
RESULT   7
CNS0181E/c
LOCUS       CNS0181E     1101 bp    DNA              GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108764
VERSION     AL108764.1  GI:5629068
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37H05"
                     /note="end : SP6"
BASE COUNT       61 a    215 c    301 g    235 t    289 others
ORIGIN
```

```
  Query Match             3.7%;  Score 46.6;  DB 83;  Length 1101;
  Best Local Similarity   40.2%;  Pred. No. 6.4;
  Matches  101;  Conservative   40;  Mismatches  109;  Indels    1;  Gaps    1;

Qy     410 CCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGC 469
           ||  : :::::  ::  :||   :|: ::  || : |::||| ||:| :||  :||: :
Db     308 CCCCMCVVVMVVGMMAASCCCCAVASCSSCCCCSSCCSSCCGCCGSCCSACSCSCCSCSG 249

Qy     470 TTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGA 529
             |:| :   : ||  |   || |     | |:|| |:|  :   ||:|   :| :: :
Db     248 AAGMCGSMSGVGGCACAAGGGCGGAGACCCGCSCCACSCCCSSSCGCSCSSCSCCSSSSS 189

Qy     530 TGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCG 589
            |||| |||   |||||| :|  |  |     | |:|   :  | ||  |  :::|| :|
Db     188 CGCGACCCCSSAGGGGCCCSGCCGCCCACCCCGCCSGACCVGACCGCCAC-CMMVGGGSG 130

Qy     590 ACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCA 649
           |||  :  || ||  ||  | |   || |  | ||  | | ||||||  |||    ||||
Db     129 ACGCMAAGGGGAGAGGGCCGGNAGCCCCGCNCCACCCCCCCGCCCGCCACGCGCCGGGCA 70

Qy     650 TGGGCGCGACC 660
           | || | |||
Db      69 AGCGCCCAACC 59
```

```
RESULT   8
CNS006XK
LOCUS       CNS006XK     935 bp    DNA              GSS       03-JUN-1999
DEFINITION  Dro: ophila melanogaster genome survey sequence T7 end of BAC #
            BAC! '4N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N09"
                     /note="end : T7"
BASE COUNT       257 a    170 c    162 g     96 t    250 others
```

```
ORIGIN

  Query Match             3.7%;  Score 46.4;  DB 82;  Length 935;
  Best Local Similarity   27.7%;  Pred. No. 6.9;
  Matches   77;  Conservative   76;  Mismatches  125;  Indels    0;  Gaps    0;

Qy    389 GCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCG 448
          :| :|:  : :|   |  :: ::::: |: :||| : ::|| |::  |   |  |  ||
Db    511 SCSMCRTSSGSTGCSTCCMYMSSSVSCCSCSGTCCGYCSSCTSCKMCSCTYGCKCGCGCG 570

Qy    449 CCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCC 508
          |:  |::::| :: :::  :||:   |::  : |:|  | ||  |: || | :||   |
Db    571 CSTSCSSSSCCSBBSYSTCSCTBCTKCSSGCSTGSGCTGCCGGGGSGCGGCGSGCGCGCG 630

Qy    509 AGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGG 568
           ||  ||  :    :  : :  ::|  | :  :|:||  : || | |  |  | |: :|
Db    631 SGGGCGGGCSSSGGSSGSSSGGSSGSSCGSSSSGSGGGGSGGGCGSGGCGGCGCGSCSGC 690

Qy    569 TGCCCGGCGTCGAACCGGCCGACGTCGTGGTCATCGGCGCCGGCACCGCCGGCTACAACG 628
          :| |::   :|  :  |: | || || :  :  ||:  ::|:: |: |:      :  ::
Db    691 GSCGCSSGCGSGCGSGGGSGGGCGSCGCSCSSCGCGSGCSSGSSGCSCCSCCGSCSCSSS 750

Qy    629 CAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACG 666
          : :||:    ||::: :|    : |  |: |||   ||
Db    751 SGSCCSCGCGCGMSMCSGCGGSSCGGCCSGCCGCSSCG 788


RESULT   9
CNS012I3/c
LOCUS       CNS012I3      645 bp    DNA              GSS       ?-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7   d of BAC
            BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL101589
VERSION     AL101589.1  GI:5613200
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN08C07"
                     /note="end : T7"
BASE COUNT       28 a     26 c     85 g     92 t    414 others
ORIGIN


  Query Match             3.7%;  Score 46.2;  DB 82;  Length 645;
  Best Local Similarity   10.3%;  Pred. No. 7.1;
  Matches   43;  Conservative  135;  Mismatches  238;  Indels    3;  Gaps    1;

Qy    332 GCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCA 391
          :::  :    |::   | |      ||    || ||   |||::: ::: :    ::  ::
Db    633 SSSNSNNNNASSNNNANANANNNTTTTTTTTNTTTTTTTTSSSNSSSSSCAAASSSNSSN 574

Qy    392 CCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCG 451
          | |   :   :   :    |:::||  :: :|   : |  :     :   :    : :::
Db    573 CAGSNNSCNGSNNNSSNSNCSSVACAGSSSSANCSSACSCSNSNNNSNNNSSNNNSNSSS 514

Qy    452 ACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGG 511
          |     |   |   :   |:    :  :  ::||  ::   : :    :::: |    :
Db    513 ATNNNNCTANAANASSANNTSAANGSSNSNNSSAAANSSNAASASNTTNSSSSNCNNNNS 454

Qy    512 TTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGC 571
            :: :  |  |:: :  : :  ::  |:   :::: :   :  :   :: :  :  :
Db    453 NNNSSTSSNAAAASSSSCASNSSSSSASASNAASSSSAASSANSNSSNSASSASANSSAS 394

Qy    572 CCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG---CGCCGGCACCGCCGGCTACAACG 628
          :::::: ::  :|  :: | : :  :   :  :       ::: :::::::| : | |:
Db    393 SSSSSSSSSCSSCNASSAATSNSNNSNSAVSNNSTNNNNNNSSSNSSSSSSSCSVAACCS 334

Qy    629 CAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACA 688
          :  ::::|||:  : |: :    : ::::  :     ::      :       |      :|
Db    333 SSNSSSSCATSNNSSASNSANAASASSSSNNSNNSNASSNNNNSSNNNSNANNSNNNSSA 274

Qy    689 AACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCC 747
            :::|    :   :::      :::|:::          :       |   : :
Db    273 NSNSSSSSCSCNNSNSNNSSSCCSNNNNSSSCSSSNNNNSNNNSNNSNNNNNANNNSNS 215


RESULT  10
CNS006XK/c
LOCUS       CNS006XK      935 bp    DNA              GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N09"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
ORIGIN
```

```
  Query Match             3.6%;  Score 45.8;  DB 82;  Length 935;
  Best Local Similarity   27.7%;  Pred. No. 9;
  Matches   94;  Conservative   83;  Mismatches  162;  Indels    0;  Gaps    0;

Qy    428 CCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAG 487
          |:  || : ::: |    : |:|| | |:   :  ||::  |  ||| :   | ||
Db    856 CSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGSCSCCSSSCCGCGCCGSCSCGCGCCGGC 797

Qy    488 TCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCC 547
            |||:| || : || |:  | |:   ||: :  |   | :  :|:  : :   ||:| :
Db    796 SGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKSKCGCGCGSGGSCSSSSGSGSCGGSGGS 737

Qy    548 GCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCG 607
          ||:: : ||    |:: :| |   || :|| :   :|| ::|  :    |: |  | ||
Db    736 GCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSCGCSCGCSSGCGSCGCSGSCGCGCCGCC 677

Qy    608 CCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGG 667
          :|| | :| ||: :  :  ::| :||::  : |:     |    :|||| |:   ||
Db    676 SCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSCCSSSGCCCGCCCSCGCGCGCSCGCCGC 617

Qy    668 TTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCC 727
            | | || |     |:  :| :  :| :: : :| ::::|    |    |: :| :
Db    616 SCCCCGGCAGCSCASGCSSGMAGVAGSGASRSVVSGGSSSSGSASGCGCGCGMGCRAGSG 557

Qy    728 ACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCG 766
          : :  :: :|  | :|:    :: : : ||:|  :|::
Db    556 KMGSAGSSGRCGGACSGSGGGSBSSSKRKGGASGCASCSS 518


RESULT  11
CNS017ZK
LOCUS       CNS017ZK      744 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37B10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108698
VERSION     AL108698.1  GI:5629002
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37B10"
                     /note="end : SP6"
BASE COUNT       20 a    191 c    245 g    105 t    183 others
ORIGIN


  Query Match             3.6%;  Score 45.2;  DB 83;  Length 744;
  Best Local Similarity   34.0%;  Pred. No. 11;
  Matches   87;  Conservative   41;  Mismatches  128;  Indels    0;  Gaps    0;

Qy    350 TCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATT 409
          :|:|  || | |::  ||       :| ||:| : ::::|   ||:   |       | ||
Db    191 YCYTTGTCGCYTYSGGGCYGSGCCYCYCGYCCSYYBYYTYTCCCSYGCCTCNYSCCGYTT 250

Qy    410 CCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGC 469
          :|  : :| : :   |  ||  || | |:  || | || |:  |: | | |  |||| |
Db    251 YCCTYCYCYYYYTGYTCCCCCSCGCGCCYTCCCGGCCCCCYCCCSCCCCCCCCCCCCCCC 310

Qy    470 TTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGA 529
            ||||            | ||   |  | |:||:|||          |   || ||
Db    311 CGCCCCCCCCCCCGCCCCCCCCCCCNCCCCCSCCSCCCCNNGCNNNCCNNCCCGCCNNG 370

Qy    530 TGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCG 589
          :|| :|   :||  :::::: : :: :  ||| |: | :|: |:||  |    || :|
Db    371 CSCGNGSCNGSSGGCCSSSSSSGSGSSSSGGGGGGGSCGGSCSNGSCGCGGGNGNGGGSG 430

Qy    590 ACGTCGTGGTGATCGG 605
            | | |:   | ||
Db    431 NGGGNGSGSGTKTKGG 446


RESULT  12
CNS017SY
LOCUS       CNS017SY      1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108460
VERSION     AL108460.1  GI:5628764
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37L08"
                     /note="end : SP6"
BASE COUNT      254 a    176 c    160 g    152 t    359 others
ORIGIN


  Query Match             3.6%;  Score 45.2;  DB 83;  Length 1101;
  Best Local Similarity   12.6%;  Pred. No. 12;
  Matches   40;  Conservative   157;  Mismatches  121;  Indels    0;  Gaps    0;

Qy    441 CCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACT 500
          | |:   | :|: ||:||     :|   |   :|::     :::  : : :||: ||::
Db    755 CGASGAGGGSGRGGGSGCCASSGVCGAACSSASCSSASMCGVSSGSSCSASCGSCCGVSS 814

Qy    501 CGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGAT 560
          |: :: ::|::  ::: :  |:|  : :  : ||:::: :::: :::    : :: :
Db    815 CSAVSASSASSVMSKVASAVASCSAVASGMSAGAVSSSCRSSVASSVSAAASVSSSVSSS 874
```

```
Qy     561 GGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGG 620
Db     875 SSSSVVSAAAVSASSSSASSASMAVAAAAAVAVSSVSVASVVSSSSSCSSSSAASVVVSA 934

Qy     621 CTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAA 680
Db     935 SVAASASSVSSSSSSSVSTSSSASVSSVSAVSMSAVVSSSASSSSSVSVSSVVAVAASAA 994

Qy     681 CATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTC 740
Db     995 AAAAAAAAAASSSSASAVAVVSASSSSSSASSSSSSSASSSVSSVSSSSCSVVSVVSSSS 1054

Qy     741 ATCGGCCTACGAGCTCGA 758
Db    1055 VSVAVASASASASVSSVA 1072
```

RESULT 13
AI712257
LOCUS AI712257 315 bp mRNA EST 02-FEB-2000
DEFINITION 605069D09.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION AI712257
VERSION AI712257.1 GI:5006195
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE 1 (bases 1 to 315)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246668.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605069 row: D column: 09.

FEATURES Location/Qualifiers
source 1. .315
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI; Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

BASE COUNT 61 a 105 c 106 g 42 t 1 others
ORIGIN

Query Match 3.6%; Score 45; DB 51; Length 315;
Best Local Similarity 50.7%; Pred. No. 11;
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy     412 GGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTT 471
Db      58 GGCATGGCGTCTCCATCCAAAGACAACAGCTGGGTGGCCGAGCAAACTTTAAACCACCCC 117

Qy     472 GCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATG 531
Db     118 GCCCTAAGCAGCCAAGGCGGCGGTCACGGCGATAGCGGCGGCGGCGGTCGCGAGCCACCG 177

Qy     532 CGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGAC 591
Db     178 CGTGCGGCACGGGGCCGCCGCGTTCTTCTTGTAGTCCTTCCTCTGGTCGGCCGGCGAGGC 237

Qy     592 GTCGTGGTGATCGGCGCCGGCACCGCCGGCTAC 624
Db     238 GTCGGCGGTGGCCGGGCCAGGCGCGTCCGCTCC 270
```

RESULT 14
CNS010RJ
LOCUS CNS010RJ 846 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL099337
VERSION AL099337.1 GI:5610948
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 846)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES Location/Qualifiers
source 1. .846
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xre "taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN04N13"
/note="end : T7"

BASE COUNT 300 a 47 c 40 g 175 t 284 others
ORIGIN

Query Match 3.6%; Score 45; DB 82; Length 846;
Best Local Similarity 16.2%; Pred. No. 13;
Matches 38; Conservative 109; Mismatches 88; Indels 0; Gaps 0;

```
Qy     527 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 586
Db     574 KGAGCSGRARSSAAGAGGGGGGSGSGGCGAGGSSSSSSGSSAVGSSSSSVSSSASSASSS 633

Qy     587 CCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACG 646
Db     634 SSGSSSSAAASAAGASSVSSASSAVSSSSSSASAASSSASASVSSSSSVSAVAASSASSS 693

Qy     647 GCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACG 706
Db     694 SSYSVVSSSSAVAAVAVVSASSACSASVSSSVSGCAASSASASSTSTSSSSAASSSSSSG 753

Qy     707 CCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGG 761
Db     754 SSASGASSSSAASAASSSSAAASAASSVSAVAVSASSSSASSASSSSSASSSSSS 808
```

RESULT 15
CNS0072Q
LOCUS CNS0072Q 932 bp DNA GSS 03-JUN-1999

```
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14B09"
                     /note="end : T7"
BASE COUNT      155 a    202 c    241 g     91 t    243 others
ORIGIN
```

Query Match 3.6%; Score 44.8; DB 82; Length 932;
Best Local Similarity 34.1%; Pred. No. 14;
Matches 73; Conservative 47; Mismatches 94; Indels 0; Gaps 0;

```
Qy     454 GGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTT 513
           | ::| |   ||| ::  :|||| | |:   : |||  : :| | :  |:|||  :
Db     679 GCSSCGCCSGCCCCSSGCSCCCCCGCGCGSCGGSCCGCGCSCSGGCSSCGCSCCCSCSCC 738

Qy     514 GGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCC 573
             : |   :| |: :    |   |:   |: || :||| | |: |  ||:||:::   :
Db     739 CCSCCCCSSCGCSSSSGCGGSGGCSSGCGSCGCGSCGGSGCGSCGGCGGSCGSSSGCGSG 798

Qy     574 GGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCC 633
           :: : ||  ::|::||  : |: |  :  ::|| ::|:  || |::||  |   :| :|
Db     799 SSGSGCGSSSSGSSCGGGSGCSSGCSSGGSSGCCSSGSGGCCCCSSGCGGCGCGSCSSCG 858

Qy     634 CGCATCGCCAACGGCATGGGCGCGACCGTTACGG 667
           ||:  | :   || |  || |||| | :   |||
Db     859 CGSGGCCSGSCCGCCSCGGCCGCGGCGSGSSCGG 892
```

Search completed: June 22, 2000, 12:07:07
Job time: 10341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:04:06 ; Search time 123.79 Seconds
(without alignments)
16.801 Million cell updates/sec

Title: US-09-362-485-24
Perfect score: 16
Sequence: 1 ATTTGGGTGCCTTGGC 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	14	87.5	2061	4	US-08-835-170-1	Sequence 1, Appli
c	2	14	87.5	2327	4	US-08-835-170-3	Sequence 3, Appli
	3	14	87.5	3451	4	US-08-743-637B-18	Sequence 18, Appl
	4	14	87.5	3451	5	US-08-526-840B-18	Sequence 18, Appl
c	5	14	87.5	4325	4	US-08-888-497-21	Sequence 21, Appl
c	6	14	87.5	4325	6	PCT-US94-07926-21	Sequence 21, Appl
	7	13.4	83.8	1455	2	US-08-446-803-5	Sequence 5, Appli
	8	13.4	83.8	1455	3	US-08-861-837-5	Sequence 5, Appli
c	9	13.4	83.8	1573	7	5169835-5	Patent No. 5169835
c	10	13.4	83.8	2077	1	US-07-872-644-52	Sequence 52, Appl
c	11	13.4	83.8	2077	1	US-08-297-494-52	Sequence 52, Appl
c	12	13.4	83.8	2077	1	US-08-297-510-52	Sequence 52, Appl
c	13	13.4	83.8	2077	2	US-08-479-532-52	Sequence 52, Appl
c	14	13.4	83.8	2077	2	US-08-455-526-52	Sequence 52, Appl
c	15	13.4	83.8	2077	2	US-08-455-525-52	Sequence 52, Appl
c	16	13.4	83.8	2077	5	US-09-139-491-52	Sequence 52, Appl
c	17	13.4	83.8	2077	6	PCT-US92-03222-52	Sequence 52, Appl
c	18	13.4	83.8	2693	1	US-07-872-644-50	Sequence 50, Appl
c	19	13.4	83.8	2693	1	US-08-297-494-50	Sequence 50, Appl
c	20	13.4	83.8	2693	1	US-08-297-510-50	Sequence 50, Appl
c	21	13.4	83.8	2693	2	US-08-479-532-50	Sequence 50, Appl
c	22	13.4	83.8	2693	2	US-08-455-526-50	Sequence 50, Appl
c	23	13.4	83.8	2693	2	US-08-455-525-50	Sequence 50, Appl
c	24	13.4	83.8	2693	5	US-09-139-491-50	Sequence 50, Appl
c	25	13.4	83.8	2693	6	PCT-US92-03222-50	Sequence 50, Appl
c	26	13.4	83.8	50341	2	US-08-247-901C-1	Sequence 1, Appli
c	27	13.4	83.8	50341	4	US-09-075-904-1	Sequence 1, Appli
c	28	13	81.2	1203	1	US-07-989-991A-1	Sequence 1, Appli
	29	13	81.2	2606	3	US-08-568-459A-7	Sequence 7, Appli
	30	13	81.2	2606	4	US-08-487-826B-7	Sequence 7, Appli
c	31	12.8	80.0	236	3	US-08-379-556A-5	Sequence 5, Appli
c	32	12.8	80.0	890	3	US-08-379-556A-3	Sequence 3, Appli
c	33	12.8	80.0	1211	3	US-08-379-556A-1	Sequence 1, Appli
c	34	12.8	80.0	2242	2	US-08-641-627A-37	Sequence 37, Appl
c	35	12.8	80.0	2427	3	US-08-678-039A-39	Sequence 39, Appl
	36	12.8	80.0	2689	1	US-08-465-795-2	Sequence 2, Appli
	37	12.8	80.0	5589	1	US-08-465-795-1	Sequence 1, Appli
c	38	12.8	80.0	7620	1	US-07-767-135-1	Sequence 1, Appli
c	39	12.8	80.0	7620	1	US-07-841-652-1	Sequence 1, Appli
	40	12.8	80.0	9421	3	US-08-370-319C-2	Sequence 2, Appli
c	41	12.8	80.0	12284	4	US-08-876-991-1	Sequence 1, Appli
c	42	12.8	80.0	12284	4	US-09-059-853-1	Sequence 1, Appli
	43	12.8	80.0	19932	4	US-08-477-451-25	Sequence 25, Appl
c	44	12.4	77.5	1908	1	US-08-173-508-1	Sequence 1, Appli
c	45	12.4	77.5	1908	3	US-08-265-310-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-835-170-1/c
; Sequence 1, Application US/08835170
; Patent No. 5965420
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Wei Xie
; TITLE OF INVENTION: Human Protein Kinases hYAK3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,170
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-835-170-1

Query Match 87.5%; Score 14; DB 4; Length 2061;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db    556 TTTGGGTGCCTTGG 543


RESULT   2
US-08-835-170-3/c
; Sequence 3, Application US/08835170
; Patent No. 5965420
;  GENERAL INFORMATION:
;    APPLICANT:  Creasy, Caretha
;    APPLICANT:  Wei Xie
;    TITLE OF INVENTION:  Human Protein Kinases hYAK3
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,170
;      FILING DATE:  26-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  GH50004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other
US-08-835-170-3


  Query Match             87.5%;  Score 14;  DB 4;  Length 2327;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGG 15
          ||||||||||||||
Db    822 TTTGGGTGCCTTGG 809


RESULT   3
US-08-743-637B-18
; Sequence 18, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3451 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
US-08-743-637B-18


  Query Match             87.5%;  Score 14;  DB 4;  Length 3451;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 TTGGGTGCCTTGGC 16
          ||||||||||||||
Db   1522 TTGGGTGCCTTGGC 1535


RESULT   4
US-08-526-840B-18
; Sequence 18, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  277
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3451 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas aeruginosa
US-08-526-840B-18


  Query Match             87.5%;  Score 14;  DB 5;  Length 3451;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 TTGGGTGCCTTGGC 16
           ||||||||||||||
Db    1522 TTGGGTGCCTTGGC 1535


RESULT   5
US-08-888-497-21/c
; Sequence 21, Application US/08888497
; Patent No. 5972677
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,405
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2409
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  722..1195
US-08-888-497-21


  Query Match             87.5%;  Score 14;  DB 4;  Length 4325;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGG 15
           ||||||||||||||
Db    1930 TTTGGGTGCCTTGG 1917


RESULT   6
PCT-US94-07926-21/c
; Sequence 21, Application PC/TUS9407926
;  GENERAL INFORMATION:
;    APPLICANT:  Tischfield, Jay A.
;    APPLICANT:  Seilhamer, Jeffrey J.
;    TITLE OF INVENTION:  Mammalian Phospholipase A2 Nucleotide
;    TITLE OF INVENTION:  Sequences and Low Molecular Weight Amino Acid Sequences
;    TITLE OF INVENTION:  Encoded Thereby, Antisense Sequences and Nucleotide
;    TITLE OF INVENTION:  Sequences Having Internal Ribosome Binding Sites
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ruden, Barnett, McClosky, Smith, Schuster &
;      ADDRESSEE:  Russell PA
;      STREET:  200 East Broward Boulevard
;      CITY:  Fort Lauderdale
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  33301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07926
;      FILING DATE:  15-JUL-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/097,354
;      FILING DATE:  26-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Manso, Peter J.
;      REGISTRATION NUMBER:  32,264
;      REFERENCE/DOCKET NUMBER:  IN21044-5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  305-527-2498
;      TELEFAX:  305-764-4996
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  722..1195
```

PCT-US94-07926-21

Query Match 87.5%; Score 14; DB 6; Length 4325;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGG 15
||||||||||||||
Db 1930 TTTGGGTGCCTTGG 1917

RESULT 7
US-08-446-803-5
; Sequence 5, Application US/08446803
; Patent No. 5824531
; GENERAL INFORMATION:
; APPLICANT: Ottrup, Helle
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Rasmussen, Michael Dolberg
; APPLICANT: Van Der Zee, Pia
; TITLE OF INVENTION: Alkaline Bacillus Amylase
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5824531o No. 5824531disk of No. 5824531th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,803
; FILING DATE: 01-June-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4157.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-446-803-5

Query Match 83.8%; Score 13.4; DB 2; Length 1455;
Best Local Similarity 93.3%; Pred. No. 76;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
|||| ||||||||||
Db 812 ATTTAGGTGCCTTGG 826

RESULT 8
US-08-861-837-5
; Sequence 5, Application US/08861837
; Patent No. 5856164
; GENERAL INFORMATION:
; APPLICANT: Ottrup, Helle
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Rasmussen, Michael Dolberg
; APPLICANT: Van Der Zee, Pia
; TITLE OF INVENTION: Alkaline Bacillus Amylase
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5856164o No. 5856164disk of No. 5856164th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,837
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,803
; FILING DATE: 01-June-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4157.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-861-837-5

Query Match 83.8%; Score 13.4; DB 3; Length 1455;
Best Local Similarity 93.3%; Pred. No. 76;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
|||| ||||||||||
Db 812 ATTTAGGTGCCTTGG 826

RESULT 9
5169835-5/c
;Patent No. 5169835
; APPLICANT: WAI-YEE, CHAN
; TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
; NUMBER OF SEQUENCES: 48
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/390,409
; FILING DATE: 07-AUG-1989
;SEQ ID NO:5:
; LENGTH: 1573
5169835-5

Query Match 83.8%; Score 13.4; DB 7; Length 1573;
Best Local Similarity 93.3%; Pred. No. 77;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
|||||||| ||||||
Db 177 TTTGGGTGGCTTGGC 163

```
RESULT  10
US-07-872-644-52/c
; Sequence 52, Application US/07872644
; Patent No. 5389527
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,644
;      FILING DATE:  19920420
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5389527and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30822
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2077 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..1693
US-07-872-644-52


  Query Match             83.8%;  Score 13.4;  DB 1;  Length 2077;
  Best Local Similarity   93.3%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGGC 16
           ||||||| |||||||
Db    1147 TTTGGGTACCTTGGC 1133


RESULT  11
US-08-297-494-52/c
; Sequence 52, Application US/08297494
; Patent No. 5580771
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/297,494
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5580771and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30822
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2077 base pairs
;      TYPE:  n cleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..1693
US-08-297-494-52


  Query Match             83.8%;  Score 13.4;  DB 1;  Length 2077;
  Best Local Similarity   93.3%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGGC 16
           ||||||| |||||||
Db    1147 TTTGGGTACCTTGGC 1133


RESULT  12
US-08-297-510-52/c
; Sequence 52, Application US/08297510
; Patent No. 5602019
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
```

```
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/297,510
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5602019and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30822
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2077 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..1693
US-08-297-510-52


  Query Match             83.8%;  Score 13.4;  DB 1;  Length 2077;
  Best Local Similarity   93.3%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGGC 16
           ||||||| |||||||
Db    1147 TTTGGGTACCTTGGC 1133


RESULT  13
US-08-479-532-52/c
; Sequence 52, Application US/08479532
; Patent No. 5776752
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,532
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/297,494
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5776752and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30822
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2077 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..1693
US-08-479-532-52


  Query Match             83.8%;  Score 13.4;  DB 2;  Length 2077;
  Best Local Similarity   93.3%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGGC 16
           ||||||| |||||||
Db    1147 TTTGGGTACCTTGGC 1133


RESULT  14
US-08-455-526-52/c
; Sequence 52, Application US/08455526
; Patent No. 5789553
;  GENERAL INFORMATION:
;    APPLICANT:  Beavo, Joseph A.
;    APPLICANT:  Bentley, Kelley
;    APPLICANT:  Charbonneau, Harry
;    APPLICANT:  Sonnenburg, William K.
;    TITLE OF INVENTION:  DNA Encoding Mammalian
;    TITLE OF INVENTION:  Phosphodiesterases
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,526
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/297,494
;      FILING DATE:  29-AUG-1994
;      APPLICATION NUMBER:  US 07/688,356
;      FILING DATE:  04-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5789553and, Greta E.
;      REGISTRATION NUMBER:  35,302
```

; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
US-08-455-526-52

Query Match 83.8%; Score 13.4; DB 2; Length 2077;
Best Local Similarity 93.3%; Pred. No. 80;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
||||||| |||||||
Db 1147 TTTGGGTACCTTGGC 1133

RESULT 15
US-08-455-525-52/c
; Sequence 52, Application US/08455525
; Patent No. 5800987
; GENERAL INFORMATION:
; APPLICANT: Beavo, Joseph A.
; APPLICANT: Bentley, Kelley
; APPLICANT: Charbonneau, Harry
; APPLICANT: Sonnenburg, William K.
; TITLE OF INVENTION: DNA Encoding Mammalian
; TITLE OF INVENTION: Phosphodiesterases
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,525
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/297,494
; FILING DATE:
; APPLICATION NUMBER: US 07/688,356
; FILING DATE: 04-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5800987and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30822
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1693
US-08-455-525-52

Query Match 83.8%; Score 13.4; DB 2; Length 2077;
Best Local Similarity 93.3%; Pred. No. 80;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
||||||| |||||||
Db 1147 TTTGGGTACCTTGGC 1133

Search completed: June 22, 2000, 18:04:08
Job time: 9715 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:38 ; Search time 2113.77 Seconds
(without alignments)
30.681 Million cell updates/sec

Title: us-09-362-485-24
Perfect score: 16
Sequence: 1 ATTTGGGTGCCTTGGC 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	15	93.8	238	20	Z38152	Z38152 HSC01B102 n
	2	15	93.8	245	60	AV172031	AV172031 AV172031
	3	15	93.8	311	47	AI495572	AI495572 sa99f05.y
c	4	15	93.8	334	44	AI291819	AI291819 qm80f09.x
	5	15	93.8	337	21	F11038	F11038 HSC37D112 n
	6	15	93.8	398	45	AI375891	AI375891 tc14a09.x
	7	15	93.8	399	45	AI372879	AI372879 qz92e05.x
	8	15	93.8	458	79	AW306550	AW306550 se52d01.y
c	9	15	93.8	469	79	AW259655	AW259655 up31f06.x
	10	15	93.8	513	61	AI855484	AI855484 sc16f10.y
c	11	15	93.8	616	80	AW350682	AW350682 GM210009B
	12	15	93.8	663	108	AQ585628	AQ585628 RPCI-11-4
c	13	15	93.8	674	80	AW350096	AW350096 GM210007B
	14	15	93.8	697	103	AQ194038	AQ194038 RPCI11-47
c	15	15	93.8	831	85	AQ749517	AQ749517 HS_5574_A
c	16	14.4	90.0	208	28	AA108148	AA108148 ml58h06.r
	17	14.4	90.0	242	71	AV261462	AV261462 AV261462
	18	14.4	90.0	257	27	AA023751	AA023751 mh81f01.r
	19	14.4	90.0	270	30	AA274532	AA274532 va91h10.r
	20	14.4	90.0	310	61	AI818776	AI818776 wl11f09.x
c	21	14.4	90.0	322	106	AQ398026	AQ398026 mgxb0019J
c	22	14.4	90.0	324	106	AQ397965	AQ397965 mgxb0019J
c	23	14.4	90.0	342	81	AW424844	AW424844 660039B07
	24	14.4	90.0	346	94	AQ038909	AQ038909 CIT-HSP-2
	25	14.4	90.0	360	46	AI429174	AI429174 ml54a11.x
	26	14.4	90.0	361	41	AI071811	AI071811 UI-R-C2-n
	27	14.4	90.0	361	79	AW259023	AW259023 um76d10.y
	28	14.4	90.0	362	38	AA752560	AA752560 96SU0047
	29	14.4	90.0	363	47	AI527787	AI527787 uj29b04.y
c	30	14.4	90.0	378	43	AI204643	AI204643 qd44a04.x
	31	14.4	90.0	379	28	AA105433	AA105433 ml83g01.r
	32	14.4	90.0	385	63	AI982550	AI982550 wt21h03.x
	33	14.4	90.0	390	94	AQ038907	AQ038907 CIT-HSP-2
	34	14.4	90.0	403	59	AI763980	AI763980 UI-R-Y0-a
c	35	14.4	90.0	414	50	AI687689	AI687689 tp93a11.x
c	36	14.4	90.0	433	102	AQ141004	AQ141004 HS_3141_B
c	37	14.4	90.0	441	50	AI678755	AI678755 tu69f06.x
c	38	14.4	90.0	452	85	AQ698622	AQ698622 HS_5564_B
c	39	14.4	90.0	466	38	AA775413	AA775413 ad13c06.s
	40	14.4	90.0	472	23	R93559	R93559 yq34c12.r1
c	41	14.4	90.0	475	34	AA513649	AA513649 nh79f11.s
c	42	14.4	90.0	485	81	AW425443	AW425443 57408 MAR
	43	14.4	90.0	488	74	AW172031	AW172031 618047G06
c	44	14.4	90.0	494	41	AI013821	AI013821 EST208496
	45	14.4	90.0	506	103	AQ177267	AQ177267 HS_2192_A

ALIGNMENTS

```
RESULT   1
Z38152
LOCUS       Z38152         238 bp    mRNA            EST       27-OCT-1994
DEFINITION  HSC01B102 normalized infant brain cDNA Homo sapiens cDNA clone
            c-01b10 3', mRNA sequence.
ACCESSION   Z38152
VERSION     Z38152.1  GI:560160
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read. 16 T removed at sequence 5'end
            Genexpress_library_idt: C; Genexpress_sequence_idt: a1c-01b10
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="c-01b10"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total
                     brain;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       58 a     50 c     54 g     76 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 20;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   15;  onservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATTTGGGTGCCTTGG 15
           |||||||||||||||
Db     184 ATTTGGGTGCCTTGG 198


RESULT   2
AV172031
LOCUS       AV172031       245 bp    mRNA            EST       06-JUL-1999
DEFINITION  AV172031 Mus musculus head C57BL/6J 17-day embryo Mus musculus cDNA
            clone 3300003E14, mRNA sequence.
ACCESSION   AV172031
VERSION     AV172031.1  GI:5378469
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1133462.
            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
```

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .245
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3300003E14"
/clone_lib="Mus musculus head C57BL/6J 17-day embryo"
/sex="mixed"
/tissue_type="head"
/dev_stage="17-day embryo"

BASE COUNT 50 a 72 c 53 g 70 t
ORIGIN

Query Match 93.8%; Score 15; DB 60; Length 245;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
||||||||||||||
Db 136 TTTGGGTGCCTTGGC 150

RESULT 3
AI495572
LOCUS AI495572 311 bp mRNA EST 01-DEC-1999
DEFINITION sa99f05.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1004-7450 5' similar to SW:IN37_SPIOL P23525 37 KD CHLOROPLAST INNER ENVELOPE MEMBRANE PROTEIN PRECURSOR ;, mRNA sequence.
ACCESSION AI495572
VERSION AI495572.1 GI:4396575
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 311)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187734.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 300.

FEATURES Location/Qualifiers
source 1. .311
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1004-7450"
/clone_lib="Gm-c1004"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First- strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@nau.edu, virginia.coryell@nau.edu"

BASE COUNT 73 a 58 c 85 g 95 t
ORIGIN

Query Match 93.8%; Score 15; DB 47; Length 311;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
||||||||||||||
Db 217 TTTGGGTGCCTTGGC 231

RESULT 4
AI291819/c
LOCUS AI291819 334 bp mRNA EST 29-JAN-1999
DEFINITION qm80f09.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1895081 3', mRNA sequence.
ACCESSION AI291819
VERSION AI291819.1 GI:3934593
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797140.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 509 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 294.

FEATURES Location/Qualifiers
source 1. .334
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="IMAGE:1895081"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 82 a 99 c 59 g 94 t
ORIGIN

Query Match 93.8%; Score 15; DB 44; Length 334;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
||||||||||||||||
Db 327 ATTTGGGTGCCTTGG 313

RESULT 5
F11038
LOCUS F11038 337 bp mRNA EST 13-MAR-1995
DEFINITION HSC37D112 normalized infant brain cDNA Homo sapiens cDNA clone c-37d11 3', mRNA sequence.
ACCESSION F11038
VERSION F11038.1 GI:706584
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. removed at sequence 5'end
Genexpress_library_idt: C; Genexpress_sequence_idt: a2c-37d11
Insert Length: 1205 Std Error: 0.00
Seq primer: (-21)M13_universal
High quality sequence stop: 392.

FEATURES Location/Qualifiers
source 1. .337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-37d11"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII; Site_2: NotI; sex=Female; dev_stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Soares, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"

BASE COUNT 90 a 72 c 66 g 106 t 3 others
ORIGIN

Query Match 93.8%; Score 15; DB 21; Length 337;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
||||||||||||||||
Db 184 ATTTGGGTGCCTTGG 198

RESULT 6
AI375891
LOCUS AI375891 398 bp mRNA EST 18-MAR-1999
DEFINITION tc14a09.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:2063800 3', mRNA sequence.
ACCESSION AI375891
VERSION AI375891.1 GI:4175881
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 3, 1998 this sequence version replaced gi:3018943.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-15[illegible]
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 507 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2063800"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 103 a 94 c 77 g 124 t
ORIGIN

Query Match 93.8%; Score 15; DB 45; Length 398;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
||||||||||||||||

```
Db    184 ATTTGGGTGCCTTGG 198


RESULT   7
AI372879
LOCUS       AI372879      399 bp    mRNA            EST       16-FEB-1999
DEFINITION  qz92e05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:2042048 3', mRNA sequence.
ACCESSION   AI372879
VERSION     AI372879.1  GI:4152745
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 1, 1998 this sequence version replaced gi:3105843.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 503    Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2042048"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      104 a     94 c     77 g    124 t
ORIGIN


  Query Match           93.8%;  Score 15;  DB 45;  Length 399;
  Best Local Similarity  100.0%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGG 15
          |||||||||||||||
Db    184 ATTTGGGTGCCTTGG 198


RESULT   8
AW306550
LOCUS       AW306550      458 bp    mRNA            EST       20-JAN-2000
DEFINITION  se52d01.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1017-2522 5' similar to SW:IN37_SPIOL P23525 37 KD INNER
            ENVELOPE MEMBRANE PROTEIN, CHLOROPLAST PRECURSOR ;, mRNA sequence.
ACCESSION   AW306550
VERSION     AW306550.1  GI:6718903
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales, Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5406990.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-2522"
                     /clone_lib="Gm-c1017"
                     /tissue_type="vegetable buds of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from vegatable buds of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      106 a     84 c    118 g    150 t
ORIGIN


  Query Match           93.8%;  Score 15;  DB 79;  Length 458;
  Best Local Similarity  100.0%;  Pred. No. 4.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGGC 16
          |||||||||||||||
Db    320 TTTGGGTGCCTTGGC 334


RESULT   9
AW259655/c
LOCUS       AW259655      469 bp    mRNA            EST       23-DEC-1999
DEFINITION  up31f06.x1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:2655971 3',
            mRNA sequence.
ACCESSION   AW259655
VERSION     AW259655.1  GI:6632636
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 469)
```

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5422081.
            Other_ESTs: up31f06.y1
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llni.gov/bbrp/image/image.html

            MGI:1034119
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .469
                     /organism="Mus musculus"
                     /strain="FVB-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2655971"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      119 a    136 c     88 g    126 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 79;  Length 469;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATTTGGGTGCCTTGG 15
          |||||||||||||||
Db    290 ATTTGGGTGCCTTGG 276


RESULT  10
AI855484
LOCUS       AI855484       513 bp    mRNA            EST       06-DEC-1999
DEFINITION  sc16f10.y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1013-836 5' similar to SW:IN37_SPIOL P23525 37 KD CHLOROPLAST
            INNER ENVELOPE MEMBRANE PROTEIN PRECURSOR ;, mRNA sequence.
ACCESSION   AI855484
VERSION     AI855484.1  GI:5508926
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales, Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138329.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 393.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1013-836"
                     /clone_lib="Gm-c1013"
                     /tissue_type="Whole seedlings, 2-3 week old seedlings,
                     greenhouse grown"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from whole seedlings of 2-3 week old greenhouse grown
                     plants. The cDNA library was prepared using the Stratagene
                     pBluescript II XR cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly (dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into XL10-Gold host cells. This library
                     was constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      126 a     94 c    130 g    162 t      1 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 61;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 TTTGGGTGCCTTGGC 16
          |||||||||||||||
Db    341 TTTGGGTGCCTTGGC 355


RESULT  11
AW350682/c
LOCUS       AW350682       616 bp    mRNA            EST       01-FEB-2000
DEFINITION  GM210009B10B5R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION   AW350682
VERSION     AW350682.1  GI:6848392
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     On Nov 22, 1999 this sequence version replaced gi:6461744.
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
```

```
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Glycine max"
                      /cultivar="Williams"
                      /db_xref="taxon:3847"
                      /clone_lib="Gm-r1021"
                      /tissue_type="root"
                      /lab_host="XL10-Gold"
                      /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                      XhoI; Library Gm-r1021 is a sequence-driven, reracked set
                      of the original  library Gm-c1004 which was prepared from
                      root cDNA. The mRNA was isolated  from entire roots of 8
                      day old 'Williams' seedlings which were propagated  on
                      paper towels with distilled water. Stratagene's cDNA
                      Synthesis  Kit (catalog #200401) was used to synthesize
                      the cDNA. The Gm-c1004  library was constructed by Dr.
                      Paul Keim & Virginia H. Coryell,  Department of Biology,
                      Box5640, Northern Arizona University, Flagstaff,  AZ
                      86011, email: paul.keim@nau.edu, virginia.coryell@nau.edu.
                      The  contig analysis to select unique genes was performed
                      by the laboratory  of Ernest Retzel, Computational Biology
                      Centers, University of Minnesota,
                      http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                      . Reracking  was performed by Genome Systems, St. Louis,
                      http://www.genomesystems.com,  and sequencing by the Keck
                      Center for Comparative and Functional  Genomics,
                      University of Illinois,
                      http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT      206 a    141 c    106 g    160 t      3 others
ORIGIN


  Query Match             93.8%;  Score 15;  DB 80;  Length 616;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 TTTGGGTGCCTTGGC 16
           |||||||||||||||
Db     396 TTTGGGTGCCTTGGC 382


RESULT  12
AQ585628
LOCUS       AQ585628      663 bp    DNA             GSS       07-JUN-1999
DEFINITION  RPCI-11-457G19.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-457G19, genomic survey sequence.
ACCESSION   AQ585628
VERSION     AQ585628.1  GI:5012674
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .663
                      /organism="Homo sapiens"
                      /db_xref="GDB:7675266"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-457G19"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      193 a    106 c    116 g    248 t
ORIGIN


  Query Match             93.8%;  Score 15;  DB 108;  Length 663;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATTTGGGTGCCTTGG 15
             ||||||| ||||||||
Db     202 ATTTGG( TGCCTTGG 216


RESULT  13
AW350096/c
LOCUS       AW350096      674 bp    mRNA            EST       01-FEB-2000
DEFINITION  GM210007B10C8R Gm-r1021 Glycine max cDNA 3', mRNA sequence.
ACCESSION   AW350096
VERSION     AW350096.1  GI:6847806
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1135868.
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES              Location/Qualifiers
     source           1. .674
                      /organism="Glycine max"
```

/cultivar="Williams"
/db_xref="taxon:3847"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library Gm-r1021 is a sequence-driven, reracked set of the original library Gm-c1004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-c1004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@nau.edu, virginia.coryell@nau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT 217 a 152 c 115 g 167 t 23 others
ORIGIN

Query Match 93.8%; Score 15; DB 80; Length 674
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTGGGTGCCTTGGC 16
||||||||||||||||
Db 371 TTTGGGTGCCTTGGC 357

RESULT 14
AQ194038
LOCUS AQ194038 697 bp DNA GSS 20-APR-1999
DEFINITION RPCI11-47D24.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-47D24, genomic survey sequence.
ACCESSION AQ194038
VERSION AQ194038.1 GI:3605650
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .697
/organism="Homo sapiens"
/db_xref="GDB:7517759"
/db_xref="taxon:9606"
/clone="RPCI-11-47D24"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 191 a 119 c 130 g 257 t
ORIGIN

Query Match 93.8%; Score 15; DB 103; Length 697;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTGGGTGCCTTGG 15
|||||||||||||||
Db 449 ATTTGGGTGCCTTGG 463

RESULT 15
AQ749517/c
LOCUS AQ749517 831 bp DNA GSS 19-JUL-1999
DEFINITION HS_5574_A1_F12_SP6 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1150 Col=23 Row=K, genomic survey sequence.
ACCESSION AQ749517
VERSION AQ749517.1 GI:5536675
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 831)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas CG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3387
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1150 row: K column: 23
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 831.
FEATURES Location/Qualifiers
source 1. .831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1150 Col=23 Row=K"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 232 a 176 c 211 g 211 t 1 others
ORIGIN

Query Match 93.8%; Score 15; DB 85; Length 831;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2 TTTGGGTGCCTTGGC 16
            ||||||||||||||||
Db     717 TTTGGGTGCCTTGGC 703
```

Search completed: June 22, 2000, 17:43:42
Job time: 20092 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:47 ; Search time 783.2 Seconds
(without alignments)
-22.357 Million cell updates/sec

Title: US-09-362-485-25
Perfect score: 18
Sequence: 1 GGCGGCGAGTCGACCGGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	5	A87628	A87628 Sequence 25
c 2		18	100.0	720	5	A89759	A89759 Sequence 16
c 3		18	100.0	1194	2	MTU92472	U92472 Mycobacteri
c 4		18	100.0	1208	5	A89752	A89752 Sequence 9
c 5		18	100.0	1209	5	A87612	A87612 Sequence 9
c 6		18	100.0	1228	5	A87608	A87608 Sequence 5
c 7		18	100.0	1228	5	A89748	A89748 Sequence 5
c 8		18	100.0	1229	5	A87610	A87610 Sequence 7
c 9		18	100.0	1229	5	A89750	A89750 Sequence 7
c 10		18	100.0	1235	5	A87606	A87606 Sequence 3
c 11		18	100.0	1235	5	A87609	A87609 Sequence 6
c 12		18	100.0	1235	5	A87611	A87611 Sequence 8
c 13		18	100.0	1235	5	A89746	A89746 Sequence 3
c 14		18	100.0	1235	5	A89749	A89749 Sequence 6
c 15		18	100.0	1235	5	A89751	A89751 Sequence 8
c 16		18	100.0	1236	5	A87613	A87613 Sequence 10
c 17		18	100.0	1236	5	A89753	A89753 Sequence 10
c 18		18	100.0	1237	5	A87607	A87607 Sequence 4
c 19		18	100.0	1237	5	A89747	A89747 Sequence 4
c 20		18	100.0	1245	5	A87605	A87605 Sequence 2
c 21		18	100.0	1245	5	A89745	A89745 Sequence 2
c 22		18	100.0	1260	5	A87604	A87604 Sequence 1
c 23		18	100.0	1260	5	A89744	A89744 Sequence 1
c 24		18	100.0	2584	1	MTALADH	X63069 M.tuberculo
c 25		18	100.0	56414	1	MTV002	AL008967 Mycobacte
c 26		16.4	91.1	250	1	S76615	S76615 (promoter a
c 27		16.4	91.1	402	1	MFLAV88	X02579 M.flavescen
28		16.4	91.1	454	1	AB026682	AB026682 Edwardsie
29		16.4	91.1	1520	1	ACHBETLA	M60896 A.lyticus b
30		16.4	91.1	1520	5	E03537	E03537 DNA sequenc
31		16.4	91.1	10464	2	AE001875	AE001875 Deinococc
c 32		16	88.9	14662	1	SC6G3	AL096825 Streptomy
c 33		15.4	85.6	786	4	GGA132998	AJ132998 Gallus Ga
c 34		15.4	85.6	1059	4	GGA132999	AJ132999 Gallus Ga
c 35		15.4	85.6	1371	8	AF050286	AF050286 Ramichlor
c 36		15.4	85.6	2384	1	SRMSIK	Y08921 S.reticuli
c 37		15.4	85.6	2423	40	AF109134	AF109134 Homo sapi
c 38		15.4	85.6	6670	40	AF112980	AF112980 Homo sapi
c 39		15.4	85.6	9899	6	HSPDCM4	X80164 H.salinariu
40		15.4	85.6	10589	2	AE000509	AE000509 Escherich
c 41		15.4	85.6	10831	2	AE002069	AE002069 Deinococc
42		15.4	85.6	10929	2	AE002078	AE002078 Deinococc
43		15.4	85.6	11576	2	AE001994	AE001994 Deinococc
c 44		15.4	85.6	17973	1	MSM238027	AJ238027 Mycobacte
45		15.4	85.6	34983	2	SCM11	AL133278 Streptomy

ALIGNMENTS

```
RESULT   1
A87628
LOCUS       A87628          18 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 25 from Patent WO9836089.
ACCESSION   A87628
VERSION     A87628.1  GI:6736268
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      6 c      9 g      1 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 6.7e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db      1 GGCGGCGAGTCGACCGGC 18


RESULT   2
A89759/c
LOCUS       A89759          720 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9832862.
ACCESSION   A89759
VERSION     A89759.1  GI:6738291
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      133 a    238 c    226 g    123 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 5;  Length 720;
  Best Local Similarity  100.0%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    419 GGCGGCGAGTCGACCGGC 402


RESULT   3
MTU92472/c
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN


  Query Match           100.0%;  Score 18;  DB 2;  Length 1194;
  Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    486 GGCGGCGAGTCGACCGGC 469


RESULT   4
A89752/c
LOCUS       A89752          1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 486 GGCGGCGAGTCGACCGGC 469

RESULT 5
A87612/c
LOCUS A87612 1209 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 9 from Patent WO9836089.
ACCESSION A87612
VERSION A87612.1 GI:6736252
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1209
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 228 a 391 c 373 g 216 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1209;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 486 GGCGGCGAGTCGACCGGC 469

RESULT 6
A87608/c
LOCUS A87608 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9836089.
ACCESSION A87608
VERSION A87608.1 GI:6736248
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 486 GGCGGCGAGTCGACCGGC 469

RESULT 7
A89748/c
LOCUS A89748 1228 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 5 from Patent WO9832862.
ACCESSION A89748
VERSION A89748.1 GI:6738282
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1228)
AUTHORS Flohe,L. and Singh,M.
TITLE L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
JOURNAL Patent: WO 9832862-A 30-JUL-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1228
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1228;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 486 GGCGGCGAGTCGACCGGC 469

RESULT 8
A87610/c
LOCUS A87610 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9836089.
ACCESSION A87610
VERSION A87610.1 GI:6736250
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1229)
AUTHORS Flohe,L. and Singh,M.
TITLE TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
JOURNAL Patent: WO 9836089-A 20-AUG-1998;
FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES Location/Qualifiers
source 1. .1229
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 236 a 391 c 382 g 219 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 1229;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 485 GGCGGCGAGTCGACCGGC 468

RESULT 9
A89750/c
LOCUS A89750 1229 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9832862.

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    485 GGCGGCGAGTCGACCGGC 468


RESULT  10
A87606/c
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      ? others
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    486 GGCGGCGAGTCGACCGGC 469


RESULT  11
A87609/c
LOCUS       A87609       1235 bp    DNA             PAT       24-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    485 GGCGGCGAGTCGACCGGC 468


RESULT  12
A87611/c
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    485 GGCGGCGAGTCGACCGGC 468


RESULT  13
A89746/c
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6738280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

```
  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||||||||||
Db     486 GGCGGCGAGTCGACCGGC 469


RESULT  14
A89749/c
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||||||||||
Db     485 GGCGGCGAGTCGACCGGC 468


RESULT  15
A89751/c
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 18;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||||||||||||
Db     485 GGCGGCGAGTCGACCGGC 468
```

```
Search completed: June 22, 2000, 17:58:49
Job time: 10798 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:01:58 ; Search time 175.2 Seconds
(without alignments)
25.705 Million cell updates/sec

Title: US-09-362-485-25
Perfect score: 18
Sequence: 1 GGCGGCGAGTCGACCGGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	100.0	18	1	V49526	Mycobacterium sp.
2		18	100.0	18	1	V49624	AlaDH derived olig
3	c	18	100.0	1245	1	V49626	Mycobacterium tube
4	c	18	100.0	1260	1	V49510	Mycobacterium sp.
5	c	18	100.0	1260	1	V49625	Mycobacterium tube
6		16.4	91.1	1520	1	Q25083	Beta-lytic proteas
7	c	15.4	85.6	15079	1	Q91580	S. clavuligerus cl
8	c	15.4	85.6	44377	1	T78508	Platenolide syntha
9	c	15.4	85.6	44377	1	T80414	Platenolide syntha
10		15	83.3	1820	1	Q04731	cDNA sequence from
11		15	83.3	2615	1	Q04730	Sequence encoding
12		14.8	82.2	495	1	Q62124	Superoxide dismuta
13		14.8	82.2	543	1	V12290	Cyclophilin type P
14		14.8	82.2	1483	1	X27947	H. insolens Cel6B
15		14.8	82.2	1493	1	X27957	H. insolens Cel6B
16	c	14.8	82.2	1940	1	Q69050	Lignin peroxidase
17		14.8	82.2	2004	1	T13952	Maize-optimized VI
18		14.8	82.2	2004	1	T73997	Maize optimised-B.
19		14.8	82.2	2004	1	V16170	DNA encoding an 80
20		14.8	82.2	2010	1	Q74683	Bacillus cereus ve
21		14.8	82.2	2196	1	Q06815	Sequence encoding
22		14.8	82.2	2576	1	T13945	Maize-optimized VI
23		14.8	82.2	2576	1	T74006	Maize optimised-B.
24		14.8	82.2	2576	1	V16181	Maize optimised DN
25		14.8	82.2	2655	1	Q74682	Bacillus cereus ve
26		14.8	82.2	2655	1	T13951	Maize-optimized VI
27		14.8	82.2	2655	1	T13954	Maize-optimized VI
28		14.8	82.2	2655	1	T73996	Maize optimised-B.
29		14.8	82.2	2655	1	T74001	Maize optimised-Ba
30		14.8	82.2	2655	1	V16169	DNA encoding a 100
31		14.8	82.2	2655	1	V16174	Maize optimised DN
32		14.8	82.2	4031	1	T13947	Maize-optimized VI
33		14.8	82.2	4031	1	T74010	Maize optimised-B.
34		14.8	82.2	4031	1	V16189	Maize optimised DN
35		14.8	82.2	4260	1	N81768	Sequence encoding
36		14.8	82.2	4380	1	N80222	Sequence of Mycoba
37		14.8	82.2	4380	1	V05708	Mycobacterium tube
38	c	14.4	80.0	207	1	T39210	Cladosporium sp. 2
39	c	14.4	80.0	207	1	X05400	Cladosporium speci
40	c	14.4	80.0	5676	1	V21186	Amycolatopsis medi
41	c	14.4	80.0	8535	1	Q73731	GaLV SEATO genome.
42	c	14.4	80.0	9661	1	Q73733	Retro virus vector
43		14.4	80.0	28804	1	T37329	Sphingan biosynthe
44		14.4	80.0	28804	1	T92474	Sphingomonas genus
45		14.4	80.0	28804	1	V81474	Chromosomal fragme

ALIGNMENTS

RESULT 1
V49526
ID V49526 standard; DNA; 18 BP.
AC V49526;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH oligonucleotide AlaDH-RM.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Synthetic.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 10; 57pp; German.
CC V49512-V49526 are oligonucleotides used in a method to isolate an alanine
CC dehydrogenase (ADH) protein from a Mycobacterium sp. This protein is used
CC to diagnose tuberculosis and other mycobacterial infections (including
CC 'swimmers' disease', caused by M. marinum, a fish pathogen) in humans or
CC animals. The protein can also be used for control of epidemics and for
CC vaccination, to screen for agents with anti-mycobacterial activity, and
CC in bio-transformations that are specific for L-alanine. Also mycobacteria
CC can be identified by analysis of genomic ADH sequences. ADH is an antigen
CC that is secreted early during infection.
SQ Sequence 18 BP; 2 A; 6 C; 9 G; 1 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 1 GGCGGCGAGTCGACCGGC 18

RESULT 2
V49624
ID V49624 standard; DNA; 18 BP.
AC V49624;
DT 20-NOV-1998 (first entry)
DE AlaDH derived oligonucleotide AlaDH-RM.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium sp.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;

DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 9; Fig 2.5; 55pp; German.
CC THe Alanine dehydrogenase (AlaDH) derived oligonucleotides V49610-V49624
CC are used for diagnosing tuberculosis (TB) and other mycobacterial
CC infections in humans or animals. Kits are used for direct diagnosis of
CC TB on clinical samples (e.g. body fluids) and can differentiate between
CC pathogenic and non-virulent strains, e.g. for identifying pathogens of
CC the M.tuberculosis (M. t.) complex. The kit may also be used to
CC identify substances that inhibit mycobacteria, for combatting epidemics
CC and for vaccination follow-up. The oligonucleotides are used similarly
CC in diagnostic hybridisation tests, also for culture confirmation of
CC isolated strains and for chromosome fingerprinting to detect/
CC differentiate between mycobacteria, and for L-alanine-specific
CC biotransformation reactions. AlaDH is an early antigen, present
CC extracellularly after only a few days of growth, making it an ideal
CC drug target.
SQ Sequence 18 BP; 2 A; 6 C; 9 G; 1 T;

Query Match 100.0%; Score 18; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 1 GGCGGCGAGTCGACCGGC 18

RESULT 3
V49626/c
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 18; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 492 GGCGGCGAGTCGACCGGC 475

RESULT 4
V49510/c
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.
PF 29-JAN-1998; E00484.
PR 29-JAN-1997; EP-101339.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-427958/36.
PT Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT - used for diagnosis of tuberculosis and other mycobacterial
PT diseases, also for treatment and prevention, for drug screening and
PT for bio-transformation
PS Disclosure; Page 11; 57pp; German.
CC This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC and other mycobacterial infections (including 'swimmers' disease', caused
CC by M. marinum, a fish pathogen) in humans or animals. The protein can
CC also be used for control of epidemics and for vaccination, to screen for
CC agents with anti-mycobacterial activity, and in bio-transformations that
CC are specific for L-alanine. Also mycobacteria can be identified by
CC analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC early during infection.
SQ Sequence 1260 BP; 243 A; 403 C; 389 G; 225 T;

Query Match 100.0%; Score 18; DB 1; Length 1260;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||||||||||
Db 507 GGCGGCGAGTCGACCGGC 490

RESULT 5
V49625/c
ID V49625 standard; DNA; 1260 BP.
AC V49625;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis Alanine dehydrogenase.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Claim 13; Fig 2.3; 55pp; German.
CC THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may

```
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions. AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match            100.0%;  Score 18;  DB 1;  Length 1260;
  Best Local Similarity  100.0%;  Pred. No. 3.2;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||||||||||||||
Db    507 GGCGGCGAGTCGACCGGC 490


RESULT   6
Q25083
ID   Q25083 standard; DNA; 1520 BP.
AC   Q25083;
DT   15-NOV-1992  (first entry)
DE   Beta-lytic protease gene.
KW   Gram-negative; Gram-positive; bacteria; decomposition; ss.
OS   Achromobacter lyticus.
FH   Key             Location/Qualifiers
FT   cds             319. .1443
FT                   /*tag= a
PN   J04108387-A.
PD   09-APR-1992.
PF   29-AUG-1990; 225136.
PR   29-AUG-1990; JP-225136.
PA   (WAKP ) WAKO PURE CHEM IND.
DR   WPI; 92-171653/21.
DR   P-PSDB; R24147.
PT   Beta-lytic protease gene and DNA encoding it - for decomposing
PT   Gram-positive and some Gram-negative bacteria
PS   Claim 1; Fig 1; 13pp; Japanese.
CC   Peptidase was purified from Achromobacter lyticus by a Sepharose
CC   CL-4B column, a Sephadex G-75 column, and by reverse phase HPLC.
CC   The first 25 N-terminal amino acids were found to be identical to
CC   those of the beta-lytic protease from Lysobacter enzymogenes. The
CC   beta-lytic protease gene of A. lyticus was cloned using PCR primers
CC   based on the sequence of the L. enzymogenes beta-protease. The
CC   full length sequence of the gene encoding the A. lyticus enzyme was
CC   determined by dideoxy sequencing. The beta-lytic protease is
CC   expected to be an enzyme which can decompose not only Gram-positive
CC   bacteria but also some Gram-negative bacteria.
CC   See also Q25084-6.
SQ   Sequence   1520 BP;    258 A;    533 C;    498 G;    231 T;


  Query Match            91.1%;  Score 16.4;  DB 1;  Length 1520;
  Best Local Similarity  94.4%;  Pred. No. 16;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          ||||||| ||||||||||
Db   1243 GGCGGCCAGTCGACCGGC 1260


RESULT   7
Q91580/c
ID   Q91580 standard; DNA; 15079 BP.
AC   Q91580;
DT   13-NOV-1995  (first entry)
DE   S. clavuligerus cla gene region.
KW   Clavulinic acid; clavulinate; antibiotic; beta-lactamase-inhibitor;
KW   cla gene; ds.
OS   Streptomyces clavuligerus.
FH   Key             Location/Qualifiers
FT   cds             complement (49. .1745)
FT                   /*tag= a
FT                   /label= ORF1
FT   cds             2216. .3937
FT                   /*tag= b
FT                   /label= ORF2
FT   cds             3940. .5481
FT                   /*tag= c
FT                   /label= ORF3
FT   cds             5654. .6595
FT                   /*tag= d
FT                   /label= ORF4
FT                   /note= "cla gene"
FT   cds             6611. .7588
FT                   /*tag= e
FT                   /label= ORF5
FT   cds             7895. .9076
FT                   /*tag= f
FT                   /label= ORF6
FT   cds             9241. .10908
FT                   /*tag= g
FT                   /label= ORF7
FT   cds             complement (10998. .12296)
FT                   /*tag= h
FT                   /label= ORF8
FT   cds             complement (12662. .13365)
FT                   /*tag= i
FT                   /label= ORF9
FT   cds             13769. .14995
FT                   /*tag= j
FT                   /label= ORF10
PN   CA2108113-A.
PD   09-APR-1995.
PF   08-OCT-1993; 108113.
PR   08-OCT-1993; CA-108113.
PA   (UYAL-) UNIV ALBERTA.
PI   Aidoo KA,  Jensen SE,  Paradkar AS;
DR   WPI; 95-207301/28.
DR   P-PSDB; R77858; R77859; R77860; R77861; R77862; R77863; R77864;
DR   R77865; R77866; R77867.
PT   Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
PT   biosynthesis of the antibiotic in Streptomyces hosts which do not
PT   naturally produce clavulanate
PS   Disclosure; Fig.2; 41pp; English.
CC   A library of S. clavuligerus NRRL 5741 genomic DNA fragments (15-22
CC   kb) constructed in cosmid pLAFR3 was screened for the cla gene using
CC   a probe based on a partial N-terminal sequence from the CLA enzyme.
CC   Isolated clone K6L1 included a 15kb fragment having the sequence
CC   given in Q91580 that included the cla gene (ORF4).
SQ   Sequence   15079 BP;    2097 A;    5488 C;    5324 G;    2170 T;


  Query Match            85.6%;  Score 15.4;  DB 1;  Length 15079;
  Best Local Similarity  94.1%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGG 17
          ||||||| |||||||||
Db  13851 GGCGGCGGGTCGACCGG 13835


RESULT   8
T78508/c
ID   T78508 standard; DNA; 44377 BP.
AC   T78508;
DT   26-FEB-1998  (first entry)
DE   Platenolide synthase gene cluster.
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
OS   Streptomyces ambofaciens.
FH   Key             Location/Qualifiers
FT   CDS             350. .14002
```

```
FT                   /*tag=  a
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in W23716"
FT   CDS             14046. .20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in W23717"
FT   CDS             20110. .31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111. .20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in W23718"
FT   CDS             31329. .36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in W23719"
FT   CDS             36155. .41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in W23720"
PN   EP-791656-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301066.
PR   22-FEB-1996; US-012050.
PA   (ELIL ) LILLY & CO ELI.
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
DR   WPI; 97-418047/39.
DR   P-PSDB; W23716-W23720.
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
PS   Claim 9; Pages 8-33; 81pp; English.
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
SQ   Sequence   44377 BP;    4965 A;    15552 C;    17381 G;    6479 T;
```

```
  Query Match             85.6%;  Score 15.4;  DB 1;  Length 44377;
  Best Local Similarity   94.1%;  Pred. No. 32;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db   37671 GCGGCGAGTCGTCCGGC 37655
```

```
RESULT   9
T80414/c
ID   T80414 standard; DNA; 44377 BP.
AC   T80414;
DT   27-FEB-1998  (first entry)
DE   Platenolide synthase gene cluster.
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
OS   Streptomyces ambofaciens.
FH   Key             Location/Qualifiers
FT   CDS             350. .14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350. .352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in W22606"
FT   CDS             14046. .20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in W22607"
FT   CDS             20110. .31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111. .20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in W22608"
FT   CDS             31329. .36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in W22609"
FT   CDS             36155. .41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in W22610"
PN   EP-791655-A2.
PD   27-AUG-1997.
PF   19-FEB-1997; 301056.
PR   22-FEB-1996; US-012078.
PA   (ELIL ) LILLY & CO ELI.
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
DR   WPI; 97-418046/39.
DR   P-PSDB; W22606-W22610.
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
PS   Example 2; Pages 110-134; 220pp; English.
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see T80413) to create a hybrid ORF1 sequence (see T80415). The tylG gene
CC   is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
SQ   Sequence   44377 BP;    4965 A;    15552 C;    17381 G;    6479 T;
```

```
  Query Match             85.6%;  Score 15.4;  DB 1;  Length 44377;
  Best Local Similarity   94.1%;  Pred. No. 32;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db   37671 GCGGCGAGTCGTCCGGC 37655
```

```
RESULT  10
Q04731
ID   Q04731 standard; cDNA; 1820 BP.
AC   Q04731;
DT   12-OCT-1990  (first entry)
DE   cDNA sequence from mRNA of glucoamylase gene.
KW   Glucoamylase; fermentation; alcohol; ds.
OS   Aspergillus usamii mut. shirousamii.
PN   J02119779-A.
PD   7-MAY-1990.
PF   31-OCT-1988; 273172.
PR   31-OCT-1988; JP-273172.
PA   (JOZO-) Jozo Shigen Kenkyus.
DR   WPI; 90-182381/24.
PT   Glucoamylase gene promoter DNA sequence -
PT   isolated from Aspergillus microorganism.
PS   Disclosure; p; Japanese.
CC   Vectors constructed from the sequence may be used to transform
CC   Aspergillus oryzae or Saccharomyces cerevisiae, allowing ethanol
CC   fermentation of non-boiled rice with no other glucoamylase.
SQ   Sequence   1820 BP;   340 A;   562 C;   502 G;   416 T;
```

```
  Query Match             83.3%;  Score 15;  DB 1;  Length 1820;
  Best Local Simi arity   100.0%;  Pred. No. 68;
```

```
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GCGGCGAGTCGACCG 16
          |||||||||||||||
Db   1780 GCGGCGAGTCGACCG 1794


RESULT  11
Q04730
ID   Q04730 standard; DNA; 2615 BP.
AC   Q04730;
DT   10-MAR-1993  (revised)
DT   12-OCT-1990  (first entry)
DE   Sequence encoding glucoamylase.
KW   Glucoamylase; fermentation; alcohol; ds.
OS   Aspergillus usamii mut. shirousamii.
PN   J02119779-A.
PD   7-MAY-1990.
PF   31-OCT-1988; 273172.
PR   31-OCT-1988; JP-273172.
PA   (JOZO-) Jozo Shigen Kenkyus.
DR   WPI; 90-182381/24.
PT   Glucoamylase gene promoter DNA sequence -
PT   isolated from Aspergillus microorganism.
PS   Disclosure; p; Japanese.
CC   Vectors constructed from the sequence may be used to transform
CC   Aspergillus oryzae or Saccharomyces cerevisiae, allowing et' inol
CC   fermentation of non-boiled rice with no other glucoamylase.
SQ   Sequence   2615 BP;    521 A;    726 C;    717 G;    651 T;


  Query Match             83.3%;  Score 15;  DB 1;  Length 2615;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 GCGGCGAGTCGACCG 16
          |||||||||||||||
Db   2394 GCGGCGAGTCGACCG 2408


RESULT  12
Q62124
ID   Q62124 standard; DNA; 495 BP.
AC   Q62124;
DT   16-NOV-1994  (first entry)
DE   Superoxide dismutase gene fragment of Corynebacterium diphteriae.
KW   Superoxide dismutase; SOD; primer; probes; amplification;
KW   detection; pathogen; fungi; bacteria; mycobacteria; protozoa;
KW   infection; ss.
OS   Corynebacterium diphteriae.
PN   EP-592894-A.
PD   20-APR-1994.
PF   01-OCT-1993; 115879.
PR   13-OCT-1992; EP-810780.
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI   Zolg W;
DR   WPI; 94-127779/16.
PT   Detection and differentiation of microorganisms using the
PT   super:oxidase dismutase gene - esp for detecting pathogenic
PT   mycobacteria, using SOD specific primers for PCR amplification
PT   then genus or species specific probes
PS   Example 2; Page 28; 51pp; English.
CC   Amplification products of the superoxide dismutase (SOD) gene may be
CC   used for the differentiation of pathogenic and non-pathogenic
CC   organisms including bacteria, fungi, and protozoa.  The
CC   amplification products are generated using genus-specific primers
CC   and then tested with species specific probes.  Two primers (Q62092,
CC   Q62093) are universal primers for the SOD gene.  Q62093 may be used
CC   alongside a third primer (Q62094) for the specific amplification of
CC   the SOD gene from species of the genus Mycobacterium.  The method is
CC   particularly used to detect pathogenic mycobacteria.  This sequence
CC   corresponds to bases 188 to 666 of Corynebacterium diphteriae and was
CC   amplified using the two universal primers previously described.
SQ   Sequence   495 BP;    106 A;    136 C;    135 G;    118 T;


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 495;
  Best Local Similarity   88.9%;  Pred. No. 97;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          |||||||||  | ||||||
Db    189 GGCGGCGAGCCAACCGGC 206


RESULT  13
V12290
ID   V12290 standard; DNA; 543 BP.
AC   V12290;
DT   08-JUN-1998  (first entry)
DE   Cyclophilin type PPIase gene.
KW   Cyclophilin PPIase; halophilic; archaebacterium; immunosuppressant;
KW   cyclosporin A; ds.
OS   Halobacterium cutirubrum.
FH   Key             Location/Qualifiers
FT   CDS             1. .543
FT                   /*tag=  a
FT                   /product= "cyclophilin type PPIase"
PN   J09313184-A.
PD   09-DEC-1997.
PF   28-MAY-1996; 133353.
PR   28-MAY-1996; JP-133353.
PA   (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
DR   WPI; 98-080075/08.
DR   P-PSDB; W44191.
PT   New cyclophilin type PPIase gene  - purified from halophilic
PT   archaebacterium
PS   Claim 1; Page 5; 6pp; Japanese.
CC   The present sequence encodes a cyclophilin type PPIase gene from a
CC   halophilic archaebacterium, Halobacterium cutirubrum. The cyclophilin
CC   type PPIase may be combined with an immunosuppressant cyclosporin A.
SQ   Sequence   543 BP;    100 A;    197 C;    176 G;    70 T;


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 543;
  Best Local Similarity   88.9%;  Pred. No. 96;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGCGGCGAGTCGACCGGC 18
          |||||||||  ||||||||
Db    241 GGCGGCGACCCGACCGGC 258


RESULT  14
X27947
ID   X27947 standard; DNA; 1483 BP.
AC   X27947;
DT   08-JUN-1999  (first entry)
DE   H. insolens Cel6B fungal cellulase coding sequence.
KW   Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;
KW   cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
KW   fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
KW   detergent composition; ss.
OS   Humicola insolens.
PN   WO9901544-A1.
PD   14-JAN-1999.
PF   02-JUL-1998; DK0299.
PR   04-JUL-1997; DK-000813.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
DR   WPI; 99-106046/09.
DR   P-PSDB; Y01076.
PT   Cleaning composition containing Humicola endo-beta-1,4-glucanase  -
PT   useful as detergent compositions or additives, or as fabric
PT   conditioners
PS   Example 1; Page 246-247; 271pp; English.
```

CC This sequence encodes the Humicola insolens Cel6B fungal cellulase.
CC The invention relates to a cleaning composition (A) that contains at
CC least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
CC total cellulolytically active protein present is Humicola
CC endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
CC hydrolase family 6. (A) are used as detergent compositions or additives,
CC or as fabric softener or conditioner. (I) provides colour clarification,
CC and possibly soil removal, without significant weakening of cellulosic
CC materials during pre-soaking or wet storage (contrast use of enzymes from
CC families other than 6).
SQ Sequence 1483 BP; 285 A; 562 C; 421 G; 215 T;

Query Match 82.2%; Score 14.8; DB 1; Length 1483;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||| ||||
Db 1027 GGCGGCGAGTCGGACGGC 1044

RESULT 15
X27957
ID X27957 standard; DNA; 1493 BP.
AC X27957;
DT 08-JUN-1999 (first entry)
DE H. insolens Cel6B fungal cellulase coding sequence.
KW Cel6B; Cel6A; fungal cellulase; cleaning composition; conditioner;
KW cellulolytically active protein; endo-1,4-beta-glucanase; enzyme;
KW fabric softener; Humicola-like cellulase; glycosyl hydrolase family 6;
KW detergent composition; ss.
OS Humicola insolens.
PN WO9901544-A1.
PD 14-JAN-1999.
PF 02-JUL-1998; DK0299.
PR 04-JUL-1997; DK-000813.
PA (NOVO) NOVO-NORDISK AS.
PI Andersen KV, Damgaard B, Lund H, Nielsen JB, Schulein M;
DR WPI; 99-106046/09.
DR P-PSDB; Y01076.
PT Cleaning composition containing Humicola endo-beta-1,4-glucanase -
PT useful as detergent compositions or additives, or as fabric
PT conditioners
PS Example 3; Fig 3; 271pp; English.
CC This sequence encodes the Humicola insolens Cel6B fungal cellulase.
CC The invention relates to a cleaning composition (A) that contains at
CC least one enzyme (I) with cellulolytic activity and at least 25 wt.% of
CC total cellulolytically active protein present is Humicola
CC endo-1,4-beta-glucanase or a Humicola-like cellulase of the glycosyl
CC hydrolase family 6. (A) are used as detergent compositions or additives,
CC or as fabric softener or conditioner. (I) provides colour clarification,
CC and possibly soil removal, without significant weakening of cellulosic
CC materials during pre-soaking or wet storage (contrast use of enzymes from
CC families other than 6).
SQ Sequence 1493 BP; 289 A; 565 C; 421 G; 218 T;

Query Match 82.2%; Score 14.8; DB 1; Length 1493;
Best Local Similarity 88.9%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGCGGCGAGTCGACCGGC 18
||||||||||| ||||
Db 1063 GGCGGCGAGTCGGACGGC 1080

Search completed: June 22, 2000, 18:02:01
Job time: 10133 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:04:08 ; Search time 123.79 Seconds
(without alignments)
18.901 Million cell updates/sec

Title: US-09-362-485-25
Perfect score: 18
Sequence: 1 GGCGGCGAGTCGACCGGC 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	15.4	85.6	44377	3	US-08-804-227C-7	Sequence 7, Appli
c	2	15.4	85.6	44377	4	US-08-804-198-1	Sequence 1, Appli
	3	14.8	82.2	495	1	US-08-133-711-41	Sequence 41, Appl
	4	14.8	82.2	2004	2	US-08-471-033-18	Sequence 18, Appl
	5	14.8	82.2	2004	3	US-08-471-044-18	Sequence 18, Appl
	6	14.8	82.2	2004	3	US-08-463-483A-18	Sequence 18, Appl
	7	14.8	82.2	2004	3	US-08-471-046A-18	Sequence 18, Appl
	8	14.8	82.2	2004	3	US-08-470-566B-18	Sequence 18, Appl
	9	14.8	82.2	2004	4	US-08-469-334-18	Sequence 18, Appl
	10	14.8	82.2	2576	2	US-08-471-033-35	Sequence 35, Appl
	11	14.8	82.2	2576	3	US-08-471-044-35	Sequence 35, Appl
	12	14.8	82.2	2576	3	US-08-463-483A-35	Sequence 35, Appl
	13	14.8	82.2	2576	3	US-08-471-046A-35	Sequence 35, Appl
	14	14.8	82.2	2576	3	US-08-470-566B-35	Sequence 35, Appl
	15	14.8	82.2	2576	4	US-08-469-334-35	Sequence 35, Appl
	16	14.8	82.2	2655	2	US-08-471-033-17	Sequence 17, Appl
	17	14.8	82.2	2655	2	US-08-471-033-26	Sequence 26, Appl
	18	14.8	82.2	2655	3	US-08-471-044-17	Sequence 17, Appl
	19	14.8	82.2	2655	3	US-08-471-044-26	Sequence 26, Appl
	20	14.8	82.2	2655	3	US-08-463-483A-17	Sequence 17, Appl
	21	14.8	82.2	2655	3	US-08-463-483A-26	Sequence 26, Appl
	22	14.8	82.2	2655	3	US-08-471-046A-17	Sequence 17, Appl
	23	14.8	82.2	2655	3	US-08-471-046A-26	Sequence 26, Appl
	24	14.8	82.2	2655	3	US-08-470-566B-17	Sequence 17, Appl
	25	14.8	82.2	2655	3	US-08-470-566B-26	Sequence 26, Appl
	26	14.8	82.2	2655	4	US-08-469-334-17	Sequence 17, Appl
	27	14.8	82.2	2655	4	US-08-469-334-26	Sequence 26, Appl
	28	14.8	82.2	4031	2	US-08-471-033-49	Sequence 49, Appl
	29	14.8	82.2	4031	3	US-08-471-044-49	Sequence 49, Appl
	30	14.8	82.2	4031	3	US-08-463-483A-49	Sequence 49, Appl
	31	14.8	82.2	4031	3	US-08-471-046A-49	Sequence 49, Appl
	32	14.8	82.2	4031	3	US-08-470-566B-49	Sequence 49, Appl
	33	14.8	82.2	4031	4	US-08-469-334-49	Sequence 49, Appl
c	34	14.4	80.0	207	2	US-08-435-684A-40	Sequence 40, Appl
c	35	14.4	80.0	207	4	US-08-934-877A-40	Sequence 40, Appl
c	36	14.4	80.0	8535	5	US-08-716-351A-1	Sequence 1, Appli
c	37	14.4	80.0	9661	5	US-08-716-351A-3	Sequence 3, Appli
	38	14.4	80.0	28804	3	US-08-592-874-1	Sequence 1, Appli
	39	14.4	80.0	28804	5	US-09-096-942-2	Sequence 2, Appli
c	40	14	77.8	49377	2	US-08-764-233A-1	Sequence 1, Appli
	41	13.8	76.7	491	1	US-08-133-711-42	Sequence 42, Appl
	42	13.8	76.7	666	7	5472691-4	Patent No. 5472691
	43	13.8	76.7	669	5	US-08-556-965-1	Sequence 1, Appli
	44	13.8	76.7	723	7	5472691-7	Patent No. 5472691
	45	13.8	76.7	1155	1	US-08-457-245-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT   1
US-08-804-227C-7/c
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Fuhstoss, Stuart A.
;    APPLICANT:  Posteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7


  Query Match             85.6%;  Score 15.4;  DB 3;  Length 44377;
  Best Local Similarity   94.1%;  Pred. No. 19;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db   37671 GCGGCGAGTCGTCCGGC 37655


RESULT   2
US-08-804-198-1/c
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1


  Query Match             85.6%;  Score 15.4;  DB 4;  Length 44377;
  Best Local Similarity   94.1%;  Pred. No. 19;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 GCGGCGAGTCGACCGGC 18
           ||||||||||| |||||
Db   37671 GCGGCGAGTCGTCCGGC 37655


RESULT   3
US-08-133-711-41
; Sequence 41, Application US/08133711
; Patent No. 5525463
;  GENERAL INFORMATION:
;    APPLICANT:  Zolg, Werner
;    TITLE OF INVENTION:  Methods and reagents for detection of
;    TITLE OF INVENTION:  pathogens using superoxide dismutase gene
;    TITLE OF INVENTION:  targeting
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  NJ
;      COUNTRY:  U.S.A.
;      ZIP:  07110-1199
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/133,711
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92810780.4
;      FILING DATE:  13-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Patricia S. Rocha
;      REGISTRATION NUMBER:  31,054
;      REFERENCE/DOCKET NUMBER:  4095/95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 235 5000
;      TELEFAX:  (201) 235 3500
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  495 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Corynebacterium diphteriae/SOD gene
US-08-133-711-41


  Query Match             82.2%;  Score 14.8;  DB 1;  Length 495;
  Best Local Similarity   88.9%;  Pred. No. 59;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGCGGCGAGTCGACCGGC 18
           ||||||||| | ||||||
Db     189 GGCGGCGAGCCAACCGGC 206


RESULT   4
US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 15:18:59 ; Search time 341.15 Seconds
(without alignments)
470.941 Million cell updates/sec

Title: US-09-362-485-10
Perfect score: 1236
Sequence: 1 ATCTTGCAGATTAATCGAAC.....GGGAAGCGATGATGTCGGCC 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	47.6	3.9	1620	3	US-08-461-775-10	Sequence 10, Appl
	2	47.6	3.9	28958	1	US-08-258-261B-6	Sequence 6, Appli
	3	47.6	3.9	28958	1	US-08-456-837-6	Sequence 6, Appli
	4	47.6	3.9	28958	1	US-08-457-342-6	Sequence 6, Appli
	5	47.6	3.9	28958	1	US-08-457-646A-6	Sequence 6, Appli
	6	47.6	3.9	28958	2	US-08-458-076A-6	Sequence 6, Appli
	7	47.6	3.9	28958	2	US-08-764-233A-4	Sequence 4, Appli
	8	47.6	3.9	28958	2	US-08-457-335A-6	Sequence 6, Appli
	9	47.6	3.9	28958	2	US-08-729-214-6	Sequence 6, Appli
	10	47.6	3.9	49377	2	US-08-764-233A-1	Sequence 1, Appli
	11	46	3.7	2582	4	US-08-816-105A-2	Sequence 2, Appli
	12	45.4	3.7	833	3	US-08-403-852D-3	Sequence 3, Appli
	13	45.4	3.7	5392	3	US-08-403-852D-1	Sequence 1, Appli
	14	42.4	3.4	1833	3	US-08-403-852D-6	Sequence 6, Appli
	15	42.4	3.4	2185	1	US-08-173-508-3	Sequence 3, Appli
	16	42.4	3.4	2185	3	US-08-265-310-3	Sequence 3, Appli
	17	41.2	3.3	2414	7	5248599-1	Patent No. 5248599
	18	41.2	3.3	2668	3	US-08-461-775-11	Sequence 11, Appl
	19	41	3.3	20235	2	US-07-642-734C-3	Sequence 3, Appli
	20	40.8	3.3	44377	3	US-08-804-227C-7	Sequence 7, Appli
	21	40.8	3.3	44377	4	US-08-804-198-1	Sequence 1, Appli
c	22	40.6	3.3	4257	3	US-08-690-473-1	Sequence 1, Appli
	23	40.6	3.3	12001	2	US-08-458-568A-11	Sequence 11, Appl
	24	40.2	3.3	2064	1	US-08-343-428-1	Sequence 1, Appli
	25	40.2	3.3	11219	2	US-07-642-734C-1	Sequence 1, Appli
	26	39.6	3.2	474	3	US-08-403-852D-14	Sequence 14, Appl
	27	39.6	3.2	1524	7	5512669-1	Patent No. 5512669
	28	39.6	3.2	1525	7	5229279-1	Patent No. 5229279
	29	39.4	3.2	459	4	US-08-387-942C-35	Sequence 35, Appl
c	30	39.4	3.2	1998	7	5212296-8	Patent No. 5212296
	31	39.4	3.2	12588	4	US-08-387-942C-1	Sequence 1, Appli
c	32	39.2	3.2	420	1	US-08-470-179-148	Sequence 148, App
c	33	39.2	3.2	2353	6	PCT-US92-06840-1	Sequence 1, Appli
	34	39	3.2	43280	3	US-08-804-227C-1	Sequence 1, Appli
	35	38.8	3.1	8051	4	US-08-576-626A-2	Sequence 2, Appli
c	36	38.8	3.1	11219	2	US-07-642-734C-1	Sequence 1, Appli
	37	38.6	3.1	1187	2	US-08-440-856A-2	Sequence 2, Appli
	38	38.6	3.1	3231	2	US-08-074-121-4	Sequence 4, Appli
	39	38.6	3.1	3231	6	PCT-US94-06447-4	Sequence 4, Appli
c	40	38.4	3.1	1215	4	US-08-947-726A-1	Sequence 1, Appli
	41	38.4	3.1	2048	5	US-08-776-251-1	Sequence 1, Appli
	42	38.2	3.1	2109	4	US-08-555-568B-20	Sequence 20, Appl
c	43	38	3.1	30001	1	US-08-125-468-1	Sequence 1, Appli
c	44	38	3.1	30001	3	US-08-474-933-1	Sequence 1, Appli
	45	37.8	3.1	2588	3	US-08-796-414B-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-08-461-775-10
; Sequence 10, Application US/08461775
; Patent No. 5858773
;  GENERAL INFORMATION:
;    APPLICANT:  MAZODIER, Philippe
;    APPLICANT:  GUGLIEMI, Gerard
;    TITLE OF INVENTION:  REGULATORY NUCLEOTIDE SEQUENCE OF THE
;    TITLE OF INVENTION:  INITIATION OF TRANSCRIPTION
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  George Mason Bldg., Washington & Prince Sts.
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,775
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,313
;      FILING DATE:  10-MAY-1993
;      APPLICATION NUMBER:  FR 9011186
;      FILING DATE:  10-SEP-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Crane-Feury, Sharon E
;      REGISTRATION NUMBER:  36,113
;      REFERENCE/DOCKET NUMBER:  010830-035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1620 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1620
US-08-461-775-10
```

Query Match 3.9%; Score 47.6; DB 3; Length 1620;
Best Local Similarity 45.7%; Pred. No. 0.0031;
Matches 321; Conservative 0; Mismatches 369; Indels 12; Gaps 4;

```
Qy     114 CACCCCGGCCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGG 173
Db     879 CACCCTCACCGGTGCCACCGTCATCGCCGAGGAGGTCGGCCTCAAGCTCGACCAGGCCGG 938

Qy     174 TGCCGGAGAGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGCGCAACTGGT 233
Db     939 TCTGGACGTGCTGGGCACCGCCCGCCGCGTCACCGTCACCAAGGACGACACGACCATCGT 998

Qy     234 CGGCACCGCCGACCAGGTGTGGGCCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGAT 293
Db     999 GGACGGCGGCGGCAACGCCGAGGACGTCCAGGGCCGCGTCGCCCAGATCAAGGCCGAGAT 1058

Qy     294 AGCGGCGGAATACGGCCGCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGC 353
Db    1059 CGAGTCGACCGACTCGGACTGGGACCGCGAGAAGCTCCAGGAGCGCCTCGCCAAGCTGGC 1118

Qy     354 C-GCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACG 412
Db    1119 CGGCGGCGTCTGCGTGATCCGCGTCGGCGCGGCCACCGAGGTCGAGCTGAAGGAGCGCAA 1178

Qy     413 AGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAA--GTCGC 470
Db    1179 GCACCGTCTGGAGGACGCCATCTCCGCGACCCGCGCCGCGGTCGAGGAGGGCATCGTCTC 1238

Qy     471 CGGTCGACTCGCCGCCCAGGTTGGCGCTTACCA---CCTGATGCGAACCCAAGGGGGCCG 527
Db    1239 CGGTGGTGGCTCCGCGCTGGTCCACGCCGTCAAGGTCCTGGACGACAACCTCGGCCGCAC 1298

Qy     528 CGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG--- 585
Db    1299 CGGCGACGAGGCCACCGGTGTCGCGGTCGTCCGCCGCGCCGCCGTCGAGCCGCTGCGCTG 1358

Qy     585 ---CGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGT 641
Db    1359 GATCGCCGAGAACGCCGGCCTCGAGGGCTACGTCATCACCACCAAGGTGGCGGAGCTCGA 1418

Qy     642 TACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCG 701
Db    1419 CAAGGGCCAGGGCTTCAACGCGGCCACCGGCGAGTACGGCGACCTGGTCAAGGCCGGCGT 1478

Qy     702 GATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCT 761
Db    1479 CATCGACCCGGTCAAGGTCACCGCGTCCGCCCTGGAGAACGCGGCCTCCATCGCCTCCCT 1538

Qy     762 GGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCCAA 803
Db    1539 GCTCCTGACGACCGAGACCCTGGTCGTCGAGAAGCCGGCCGA 1580
```

RESULT 2
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match 3.9%; Score 47.6; DB 1; Length 28958;
Best Local Similarity 44.1%; Pred. No. 0.0076;
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

```
Qy     148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
Db   25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy     208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
Db   25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy     268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
Db   25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy     328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db   25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy     388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db   25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy     448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db   25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy     506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db   25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy     566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
Db   25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy     625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
```

```
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   3
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6


  Query Match             3.9%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0076;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207

Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267

Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327

Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387

Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447

Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505

Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565

Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624

Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684

Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   4
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6


  Query Match             3.9%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0076;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
          | ||| || |  ||||■|  || | ||  |   |      ||| | ||||  ||| |
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
          ||| ||      | ||| |   | | |  || |  |||| || |     |||     ||
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
          |   |  | ||| |||     |         | |   ||| ||     ||  |
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
          |   |  ||    |||| | |  |   |  || |                ||  | | |
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
            || | || || |  |   |  ||      || |  |   |||  |||| ||  | |
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
          |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | || |  |
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
            |  |||  |    | ||||  | |          ||  | || |||| |   | ||
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
          |   || |||  ||      |||||  |  || |||| |||     | |   ||||||  |
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
          | | | | |||   || |      || ||   |     |  ||       |  | |  |
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
            | | ||  |  |     |  ||   |   |  ||| ||||||||  |   ||
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
          | |   |   || |||| |      ||| || |  | |  |   ||||
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT   5
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott J eph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6


  Query Match             3.9%;  Score 47.6;  DB 1;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0076;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
          | ||| || |  ||||  ||  || | ||  |   |      ||| | ||||  ||| |
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
          ||| ||      | ||| |   | | |  || |  |||| || |     |||     ||
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425
```

```
Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT   6
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6
```

```
  Query Match             3.9%;  Score 47.6;  DB 2;  Length 28958;
  Best Local Similarity   44.1%;  Pred. No. 0.0076;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT   7
US-08-764-233A-4
```

; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1
US-08-764-233A-4

Query Match 3.9%; Score 47.6; DB 2; Length 28958;
Best Local Similarity 44.1%; Pred. No. 0.0076;
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

```
Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
          | ||| || |  |||| ||  || | ||  |    |     ||| | ||||  |||  |
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
          ||| ||      | ||| |   | | |  || |  |||| || |     |||     ||
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
          |   |  | ||| |||     |        | |    ||| ||     ||  |
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
          |   |  ||    |||| | |  |   |  || |               ||  | | |
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGCGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
          || | || || |  |   |  ||       || |  |   |||  |||| ||  | |
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
          |  | |  |  ||||    ||||||  |  | |  || | |||  | | | | || |  |
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
          |   |||   |     | ||||  | |          ||  | || |||| |   | ||
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
          |   || |||  ||      ||||| |  || |||| |||     | |   ||||||  |
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
          | | | |||   || |      || ||   |      |  ||         | | | |
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
          | | ||  |  |     |  ||   |   |    | |||  |||||||  |  ||
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
          | |   |  || |||| |      ||| || |  | |  |   ||||
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

RESULT 8
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:

```
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-335A-6
```

Query Match 3.9%; Score 47.6; DB 2; Length 28958;
Best Local Similarity 44.1%; Pred. No. 0.0076;
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

```
Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953
```

```
RESULT   9
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6
```

Query Match 3.9%; Score 47.6; DB 2; Length 28958;
Best Local Similarity 44.1%; Pred. No. 0.0076;
Matches 286; Conservative 0; Mismatches 359; Indels 3; Gaps 2;

```
Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207
Db  25306 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 25365

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267
Db  25366 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 25425

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327
Db  25426 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 25485

Qy    328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db  25486 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 25545

Qy    388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db  25546 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 25605

Qy    448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db  25606 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 25665

Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db  25666 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 25725

Qy    566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
```

```
Db  25726 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 25785

Qy    625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684

Db  25786 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 25845

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744

Db  25846 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCCGTCGAGCTCGGCGCCTCC 25905

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792

Db  25906 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 25953


RESULT 10
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologous
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs t
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS ge
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  40190..46318
;      OTHER INFORMATION:  /product= "Module 5 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  46851..47891
;      OTHER INFORMATION:  /product= "[illegible]"
;      OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;      OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;      OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;      OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1


  Query Match             3.9%;  Score 47.6;  DB 2;  Length 49377;
  Best Local Similarity   44.1%;  Pred. No. 0.0089;
  Matches  286;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    2;

Qy    148 GGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAGAGGGCTCGGCTATCACCGACGCGGAT 207

Db  41609 GCCCACGACGACCTCACCCTCGCGGATGTGGCCTATTCGCTGGCCACCACCCGCGCCCAC 41668

Qy    208 TTCAAGGCGGCAGGCGCGCAACTGGTCGGCACCGCCGACCAGGTGTGGGCCGACGCTGAT 267

Db  41669 TTCGAGCACCGCGCCGCTCTCGTAGCCCACAACCGCGACGAGCTCCTCTCCGCGCTCGAC 41728

Qy    268 TTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGCGACACGGGCAG 327

Db  41729 TCGCTCGCCCAGGACAAGCCCGCCCCGAGCACCGTCCTCGGACGGAGCGGAAGCCACGGC 41788
```

```
Qy     328 ATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGAT 387
Db   41789 AAGCTCGTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCGCTG 41848

Qy     388 TCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTG 447
Db   41849 CTCGACTCCTCGCCCGTCTTCCGCGCTCAGCTCGAAGCATGCGAGCGCGCGCTCGCTCCT 41908

Qy     448 CTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCAC--C 505
Db   41909 CACGTCGAGTGGAGCCTGCTCGCCGTCCTGCGCCGCGACGAGGGCGCCCCCTCCCTCGAC 41968

Qy     506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
Db   41969 CGCGTCGACGTCGTACAGCCCGCCCTCTTTGCCGTCATGGTCTCCCTGGCGGCCCTCTGG 42028

Qy     566 CCGACG-TCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
Db   42029 CGCTCGCTCGGCGTAGAGCCCGCCGCCGTCGTCGGCCACAGTCAGGGCGAGATCGCCGCC 42088

Qy     625 GGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
Db   42089 GCCTTCGTCGCAGGCGCTCTCTCCCTCGAGGACGCGGCCCGCATCGCCGCCCTGCGCAGC 42148

Qy     685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
Db   42149 AAAGCGCTCACCACCGTCGCCGGCAACGGGGCCATGGCCGCGTCGAGCTCGGCGCCTCC 42208

Qy     745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCC 792
Db   42209 GACCTCCAGACCTACCTCGCTCCCTGGGGCGACAGGCTCTCCATCGCC 42256


RESULT  11
US-08-816-105A-2
; Sequence 2, Application US/08816105A
; Patent No. 5989882
;  GENERAL INFORMATION:
;    APPLICANT:  Crueger, Anneliese; Dellweg, Hans-Georg; Lenz,
;    APPLICANT:  Jurgen; Schroder, Werner; Pape, Hermann;
;    APPLICANT:  Goeke, Klaus; Schaper, Beate; Hemker, Michael;
;    APPLICANT:  Piepersberg, Wolfgang; Distler, Jurgen;
;    APPLICANT:  Stratmann, Ansgar
;    TITLE OF INVENTION:  PROCESSES FOR PREPARING ACARVIOSYL
;    TITLE OF INVENTION:  TRANSFERASE AND FOR USING IT IN THE
;    TITLE OF INVENTION:  CONVERSION OF ACARBOSE HOMOLOGUES
;    TITLE OF INVENTION:  INTO ACARBOSE, FOR THE PREPARATION
;    TITLE OF INVENTION:  OF ACARBOSE HOMOLOGUES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh 6500
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect 3.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,105A
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19625269.5
;      FILING DATE:  25-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19611252.4
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9814-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2582 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-816-105A-2


  Query Match             3.7%;  Score 46;  DB 4;  Length 2582;
  Best Local Similarity   47.3%;  Pred. No. 0.009;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy     353 CCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACG 412
Db    1576 CCGCGCTGCCGGCCGGCACCTATCGCGACGTGCTCGGCGGCACCTTCGGCGGCTCCGACC 1635

Qy     413 AGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCG 472
Db    1636 TGACCGTCGAGGACGGCGACGGCACCGACCGGTCGACCGTCGCGCCGGTGCTGGGTGCCG 1695

Qy     473 GTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTG 532
Db    1696 GGCAGGTCGCCGTCTGGTCGTACCGGGCGCCGGTGGACACCGAGCCCCGGATCGGCGGGG 1755

Qy     533 TGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCA 592
Db    1756 TCGGGCCGGTCGTGACCCGGGCCGGCGCCACCGTCACCGTCGAGGGCACCGGCTTCGGCT 1815

Qy     593 CCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGG 646
Db    1816 CCGGCGGAACCGTCGCGATCGGCGGAGTCCCCGCGACCGTCCAGCAGTGGACGG 1869


RESULT  12
US-08-403-852D-3
; Sequence 3, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..833
US-08-403-852D-3


  Query Match             3.7%;  Score 45.4;  DB 3;  Length 833;
  Best Local Similarity   48.0%;  Pred. No. 0.009;
  Matches  130;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     543 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 602
           || ||  | |  || ||||    |||||||||       |       | || |  || | |
Db     327 CGAGGGCCGCACCGGCGAACTCATCGACGTCGTCCGCGGCCTGTGGGACAGCTTCGACGA 386

Qy     603 CAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACAT 662
           | ||||   |  |  || |  || |   |||  |  |     | | |  | | || ||
Db     387 CGACGCCTTCGTCCACGACCGCGCCGACGGCCTGTACTGGCGGCTGCCCGCCGTCCACCA 446

Qy     663 CGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATC 722
             | | |  | |       ||  |||  |   |||||   || ||      |  | |  |
Db     447 ACTCGACCACCAGGGCAGGCACTTCGACGTGGCCGGCCCCCTCAACGTCGCCCGCCCGCC 506

Qy     723 GGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGT 782
           |     | | ||  |  |||||||   |  ||| ||| |     |  |||| ||| ||
Db     507 GCAGGGCCACCCCGTCGTCGCCGTCACCGGCCCCGCCCTCGCCGCGGCCGCCGACCTCGT 566

Qy     783 GCCAGGCGCCAAGGCACCCAAATTAGTCTCG 813
             |   || | ||||  || |    | ||||
Db     567 CCTGCTCGACGAGGCGGCCGACGCCGCCTCG 597


RESULT  13
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03805.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
US-08-403-852D-1


  Query Match             3.7%;  Score 45.4;  DB 3;  Length 5392;
  Best Local Similarity   48.0%;  Pred. No. 0.016;
  Matches  130;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     543 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 602
           || ||  | |  || ||||    |||||||||       |       | || |  || | |
Db    3018 CGAGGGCCGCACCGGCGAACTCATCGACGTCGTCCGCGGCCTGTGGGACAGCTTCGACGA 3077

Qy     603 CAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACAT 662
           | ||||   |  |  || |  || |   |||  |  |     | | |  | | || ||
Db    3078 CGACGCCTTCGTCCACGACCGCGCCGACGGCCTGTACTGGCGGCTGCCCGCCGTCCACCA 3137

Qy     663 CGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATC 722
             | | |  | |       ||  |||  |   |||||   || ||      |  | |  |
Db    3138 ACTCGACCACCAGGGCAGGCACTTCGACGTGGCCGGCCCCCTCAACGTCGCCCGCCCGCC 3197

Qy     723 GGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGT 782
           |     | | ||  |  |||||||   |  ||| ||| |     |  |||| ||| ||
Db    3198 GCAGGGCCACCCCGTCGTCGCCGTCACCGGCCCCGCCCTCGCCGCGGCCGCCGACCTCGT 3257

Qy     783 GCCAGGCGCCAAGGCACCCAAATTAGTCTCG 813
             |   || | ||||  || |    | ||||
Db    3258 CCTGCTCGACGAGGCGGCCGACGCCGCCTCG 3288


RESULT  14
US-08-403-852D-6
; Sequence 6, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
```

```
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1689
US-08-403-852D-6


  Query Match             3.4%;  Score 42.4;  DB 3;  Length 1833;
  Best Local Similarity   50.5%;  Pred. No. 0.064;
  Matches  103;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    533 TGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCCCCGGCA 592
          || || | ||   ||||| |  || ||    | ||| |   | ||   ||| |   | |
Db    413 TGATGGTCGGGACTGGTGGTGTTCGGGCGGCGGGTCGGCCTGGGCGGCCTTCGCGCAGGACT 472

Qy    593 CCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAG 652
          |||||     || ||| |||||   || ||  ||| ||||||  | |   |   ||
Db    473 CCGCCCAACTCATCGCCGCCCGGGCCGGCATGGGCGTGGGCGGGGCGCTGCTGGCGACCA 532

Qy    653 ACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTC 712
            || |  |  || ||   | | |    ||||| |||     ||  ||||     |   |
Db    533 CCACCCTCGCCGTCATCATGCAGGTCTTCGACGACGACGAACGCCCCCGGGCGATCGGCC 592

Qy    713 GCTACTCATCGGCCTACGAGCTCG 736
```

```
                |   | |||||  |   || |
Db    593 TGTGGGGAGCGGCCAGCTCACTGG 616


RESULT  15
US-08-173-508-3
; Sequence 3, Application US/08173508
; Patent No. 5616485
;  GENERAL INFORMATION:
;    APPLICANT:  Bartfeld, Daniel
;    APPLICANT:  Butler, Michael J.
;    APPLICANT:  Hadary, Dany
;    APPLICANT:  Jenish, David
;    APPLICANT:  Krieger, Timothy
;    TITLE OF INVENTION:  STREPTOMYCES PROTEASES AND IMPROVED
;    TITLE OF INVENTION:  STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/173,508
;      FILING DATE:  23-DEC-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BEN  Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  18740/125/CACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 672 5300
;      TELEFAX:  202 672 5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2185 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  531..2069
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  531..902
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  903..2069
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  531..533
;      OTHER INFORMATION:  /note= "Met at position -124
;      OTHER INFORMATION:  represents fMet"
US-08-173-508-3


  Query Match             3.4%;  Score 42.4;  DB 1;  Length 2185;
  Best Local Similarity   47.4%;  Pred. No. 0.067;
  Matches  127;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    525 CCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG 584
          | |||  |   |||  |||| ||  |||||  ||  |   |||||  | ||  |||   |
Db   1250 CGGCGTCGGCATGACCGGCGTGGCACCCGGGGTGAAGGTGGCCGGCATCAAGGTCTCCAA 1309

Qy    585 CGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTAC 644
```

```
Db   1310 CCCCGACGGCTTCTTCTACACCGAGGCCGTGGTCTGCGGCTTCATGTGGGCGGCCGAGCA 1369

Qy    645 GGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGAT 704

Db   1370 CGGCGTCGACGTGACCAACAACAGCTATTACACCGACCCGTGGTACTTCAACTGCAAGGA 1429

Qy    705 CCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGT 764

Db   1430 CGACCCCGACCAGAAGGCGCTCGTCGAGGCCGTCTCGCGGGCCTCCCGGTACGCGGAGAA 1489

Qy    765 GATTGGGGCCGTCCTGGTGCCAGGCGCC 792

Db   1490 GAAGGGCGCGGTCAACGTCGCCGCGGCC 1517


Search completed: June 22, 2000, 15:21:26
Job time: 18764 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 12:08:02 ; Search time 5541.94 Seconds
(without alignments)
903.977 Million cell updates/sec

Title: US-09-362-485-10
Perfect score: 1236
Sequence: 1 ATCTTGCAGATTAATCGAAC.....GGGAAGCGATGATGTCGGCC 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	57.4	4.6	925	82	CNS0091P	AL053013 Drosophil
	2	55.2	4.5	925	82	CNS0091P	AL053013 Drosophil
	3	51.8	4.2	535	74	AV390505	AV390505 AV390505
c	4	50.2	4.1	844	82	CNS0052P	AL056652 Drosophil
c	5	49	4.0	910	82	CNS006ON	AL065629 Drosophil
c	6	46.6	3.8	846	82	CNS010RJ	AL099337 Drosophil
c	7	46.6	3.8	1101	83	CNS0181E	AL108764 Drosophil
	8	46.4	3.8	935	82	CNS006XK	AL066051 Drosophil
c	9	46.2	3.7	645	82	CNS012I3	AL101589 Drosophil
c	10	45.8	3.7	935	82	CNS006XK	AL066051 Drosophil
	11	45.4	3.7	591	69	AW128786	AW128786 fe37d05.y
	12	45.2	3.7	744	83	CNS017ZK	AL108698 Drosophil
	13	45.2	3.7	1101	83	CNS017SY	AL108460 Drosophil
	14	45	3.6	315	51	AI712257	AI712257 605069D09
	15	45	3.6	846	82	CNS010RJ	AL099337 Drosophil
	16	44.8	3.6	932	82	CNS0072Q	AL066742 Drosophil
c	17	44	3.6	843	82	CNS00CS1	AL059666 Drosophil
	18	43.8	3.5	427	51	AI736678	AI736678 sb32b04.y
	19	43.8	3.5	512	74	AV396794	AV396794 AV396794
	20	43.8	3.5	521	74	AV397116	AV397116 AV397116
	21	43.8	3.5	526	74	AV394827	AV394827 AV394827
c	22	43.6	3.5	1101	83	CNS016UE	AL107216 Drosophil
c	23	43.2	3.5	439	91	AQ864063	AQ864063 nbeb0022M
c	24	43.2	3.5	798	82	CNS00AJA	AL055851 Drosophil
c	25	43	3.5	413	74	AV396947	AV396947 AV396947
	26	41.8	3.4	414	80	AW286084	AW286084 LG1_261_B
	27	41.8	3.4	465	80	AW287007	AW287007 LG1_264_C
	28	41.8	3.4	530	74	AV396353	AV396353 AV396353
	29	41.8	3.4	839	82	CNS004NB	AL054280 Drosophil
c	30	41.8	3.4	932	82	CNS0072Q	AL066742 Drosophil
	31	41.6	3.4	506	74	AW180713	AW180713 MgA0863f
c	32	41.6	3.4	870	82	CNS006EZ	AL064271 Drosophil
c	33	41.4	3.3	1101	82	CNS00LXJ	AL078875 Drosophil
	34	41.4	3.3	1201	83	CNS014BJ	AL103945 Drosophil
	35	41.2	3.3	639	60	AI812194	AI812194 605086F10
	36	41.2	3.3	774	105	AQ327321	AQ327321 nbxb0040D
	37	41.2	3.3	910	82	CNS006ON	AL065629 Drosophil
	38	41	3.3	605	46	AI461529	AI461529 486018G05
c	39	41	3.3	1101	83	CNS017SY	AL108460 Drosophil
	40	40.8	3.3	384	64	AW054773	AW054773 ws59d08.x
c	41	40.8	3.3	385	21	D48746	D48746 RICS15137A
c	42	40.8	3.3	437	74	AV395790	AV395790 AV395790
	43	40.6	3.3	506	74	AV396405	AV396405 AV396405
c	44	40.4	3.3	574	91	AQ848186	AQ848186 LMAJFV1_1
c	45	40.2	3.3	916	83	CNS015SQ	AL105860 Drosophil

ALIGNMENTS

RESULT 1
CNS0091P/c
LOCUS CNS0091P 925 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 4.6%; Score 57.4; DB 82; Length 925;
Best Local Similarity 11.9%; Pred. No. 0.058;
Matches 43; Conservative 176; Mismatches 142; Indels 0; Gaps 0;

```
Qy    442 CCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTAC 501
          :|:|  : ::|:::::   ::   : ::|:::::: :|:::::::: ::  :::::   :
Db    920 SCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTS 861

Qy    502 CACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAA 561
          :||::   :::  ||     :  :|::::   |: :| |:::: :|     || :    |
Db    860 SACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSGGASA 801

Qy    562 CCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCC 621
          ::::::  :: ::  : : :: ::::: ::  ::: ||   : |::::::::::  :  :
Db    800 SHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSASAGS 741

Qy    622 AACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTC 681
          :::: :  :::|:::::|:  |::   :  : :  :| :  :: :|     :  |   |
Db    740 VVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSSASCASCSCCCTSWSC 681

Qy    682 GACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGT 741
          : : : :::    :::::::   ::::|: : :  : :  ::: :  :: :   :  ::  .
Db    680 SCSTSASMSAARSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBSMSSG 621

Qy    742 GCCGTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAGGCACCC 801
          | :| : | :| :::: :: |: :   ||||: |  : :: : : ||:|::   :: |:|
Db    620 GGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSSGCMC 561

Qy    802 A 802
          :
Db    560 R 560
```

RESULT 2
CNS0091P
LOCUS CNS0091P 925 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 4.5%; Score 55.2; DB 82; Length 925;
Best Local Similarity 11.8%; Pred. No. 0.15;
Matches 41; Conservative 170; Mismatches 137; Indels 0; Gaps 0;

```
Qy    425 CCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCG 484
          :|:|:: : :| : |  |:|:  |  :::::  :  ::  : | :::: || : :::::
Db    577 SCSCSSSSCSCCBCCGCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSSSSS 636

Qy    485 CCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCG 544
          |:: :: |::::     : :: :  :::  :::  ::: : :|:|::: |  :   |: :
Db    637 CSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTS 696

Qy    545 GGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACA 604
          ::: ::: : :: ::   :::: : :| :: :| :  :: :::::    :::::::: :
Db    697 SSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSSSYSSS 756

Qy    605 ACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCG 664
          |:|  |||:: ::::   :::  :|: :|:: ::: | ::::      :  | :|  |:
Db    757 TCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCS 816

Qy    665 ACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGG 724
          : : |:|: :|  : ::  :  : :  | |||:::    :  :   :::  : :  :::
Db    817 TYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCSSSSVS 876

Qy    725 CCTACGAGCTCGAGGGTGCCGTCAAACGTGCCGACCTGGTGATTGGGG 772
          ::  ::|:: :: ::| : ::      ::::::: :: :| |   |:::
Db    877 SSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSGSGSGSVS 924
```

RESULT 3
AV390505
LOCUS AV390505 838 bp mRNA EST 21-JAN-2000
DEFINITION AV390505 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii cDNA clone CM052b05_r, mRNA sequence.
ACCESSION AV390505
VERSION AV390505.1 GI:6544721
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 535)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE A Large Scale Structural Analysis of cDNAs in a Unicellular Green Alga, Chlamydomonas reinhardtii. I. Generation of 3433 Non-redundant Expressed Sequence Tags
JOURNAL DNA Res. 6, 369-373 (1999)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135919.
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .535
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM052b05_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 107 a 195 c 148 g 85 t
ORIGIN

Query Match 4.2%; Score 51.8; DB 74; Length 535;
Best Local Similarity 51.5%; Pred. No. 0.65;
Matches 119; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy    565 GCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAAC 624
          || ||||  ||||  ||||   || || ||  |  | |  ||||||  |  || || |
Db    221 GCTGACGAGGTGGCCATCGAGACCTGCTCCTTCTCCAAGTACGCAGGCTTCACCGGCGTG 280

Qy    625 GGCATG CGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGAC 684
          || || ||  ||||||   |      | | ||||   || |||      ||   |  ||
Db    281 CGCCTG GCTGGACCGTGGTGCCCAAGGCCCTCAAGTACGCCAACGGCGAGCCGGTGCAC 340

Qy    685 GCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCC 744
          || || |    | ||   ||   |||  ||| |      |||| | |  ||| |   ||
Db    341 GCGGACTGGAACCGCGTCATGACCACCTGCTTCAACGGCGCCTCCAACATCGTGCAGGCG 400

Qy    745 GTCAAACGTGCCGACCTGGTGATTGGGGCCGTCCTGGTGCCAGGCGCCAAG 795
          | |   |  |||  ||||  |   | ||   ||  || |     |||| |
Db    401 GGCGGTCTGGCCTGCCTGCAGCCGGAGGGGCTCAAGGAGATGAACGCCATG 451
```

RESULT 4
CNS0052P/c
LOCUS CNS0052P 844 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL056652
VERSION AL056652.1 GI:4932342
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM **Drosophila melanogaster**
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 844)
AUTHORS Genoscope.
TITLE **Direct Submission**
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR11P16"
                     /note="end : TET3"
BASE COUNT      261 a    112 c     92 g     35 t    344 others
ORIGIN
```

```
  Query Match             4.1%;  Score 50.2;  DB 82;  Length 844;
  Best Local Similarity   17.2%;  Pred. No. 1.4;
  Matches   65;  Conservative  161;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     257 CCGACGCTGATTTATTGCTCAAGGTCAAAGAACCGATAGCGGCGGAATACGGCCGCCTGC 316
           || :: ||  |:| |: | ::|   : |: :::  ::: : ::  |: ::  : :  ::
Db     768 CCMMVCCTTTTHTTTYTCMYWAMMRMYAWMMMMAMMWVTMCVMMMAVAMVMYAMAMAMRM 709

Qy     317 GACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATG 376
           : |:::| : | |:   : |:  :: :: ||   :  : |: ::| :|| :: :  |
Db     708 NNVYCSBSCNBBTTTYYCNYTCVSSSBSSSBTTCWTHTNSNTYTMVTYVTTTHH"YTNTC 649

Qy     377 CGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCC  CGGCG 436
           | :      : :| :: |  :  :  : :|:  :: : ::  :  : :: ||: | :
Db     648 CYYCCCCCCYYCSCCBYBCNTSCSYYSYSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGY 589

Qy     437 CACTAC-CCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGC 495
           | |: : |:::::|::::: :  : ::|   :|::::: :: : :::| |:| :  |  :
Db     588 CGCYSBVCSSBSSTBSSVSGSBSSTSSGTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCY 529

Qy     496 GCTTACCACCTGATGC AACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGC 555
           ::   :| ::   |::   |::   |:::::| :: |::: :  |:::||:|: :|:  |
Db     528 SSCCBYCCBSGCBTBSYBSCSSGBTGBSSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYC 469

Qy     556 GTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGC 615
           ::  ||:::::  ::|  || : |  |::::||  ::: :|::  :  ::: |:: ::
Db     468 CSBSSBCCSSSSSSTSKCCSGGSSSTKCGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSS 409

Qy     616 ATCGCCAACGGCATGGG 632
           :::::  ::::  | |
Db     408 KKYSYSTSSSSSGCGTG 392
```

```
RESULT   5
CNS006ON/c
LOCUS       CNS006ON      910 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065629
VERSION     AL065629.1  GI:4944698
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14J21"
                     /note="end : T7"
BASE COUNT      202 a     63 c    112 g    198 t    335 others
ORIGIN
```

```
  Query Match             4.0%;  Score 49;  DB 82;  Length 910;
  Best Local Similarity   20.5%;  Pred. No. 2.4;
  Matches   47;  Conservative   97;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     393 CACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGC 452
           ::::|:  : |  :::   : :|::: ::|:::::| :|::: : |: | :    :  :
Db     713 SRVMASAGSSASSSSSSSASSSASSSASSASMVVSCRSAVSVASAASAAASARSASRASA 654

Qy     453 CCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCG 512
           :: :|  |:::|:  :|:::: |: : |||: ::: :  :|:::   :: :|      :
Db     653 VMAVAVAASSSAVASSGSSSVSCSSSRCGCSASSSSASASSGSSSSSSSSSSCGCGSCAS 594

Qy     513 AACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGT 572
           |::: ::|: || :::| : :: |  :::::::: : |:::: |   :| |   |
Db     593 CASSSSRSGSCGCGSSVGCVSSSCGSSSSSSSSSSSSGCSSSSCCCCSCSCACCRCCCCC 534

Qy     573 CGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCC 621
           |:  : |  | :::||::| || || ::   :|: :| ||: |  |:||
Db     533 CRSCSCGCCCMSSSCCSSCCCCCCCCASMCCAWAMAHACCCSCCCCCDCC 485
```

```
RESULT   6
CNS010RJ/c
LOCUS       CNS010RJ      846 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099337
VERSION     AL099337.1  GI:5610948
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
```

```
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN04N13"
                     /note="end : T7"
BASE COUNT      300 a    207 c     40 g    175 t    284 others
ORIGIN


  Query Match             3.8%;  Score 46.6;  DB 82;  Length 846;
  Best Local Similarity   13.0%;  Pred. No. 6.9;
  Matches   33;  Conservative  127;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     363 TGCTTGCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCA 422
           | || :: ::   :::   :|:  ||: ::   : :: :    :::| :  : : :: : :
Db     842 TTCTSSSTSSTSCSSSSSTBTSTTTTSTSSTSASSSSSSSTSSSSSTSSTSSSSTSBTBT 783

Qy     423 GACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGC 482
           : ::  :   ::::  :| ::::| :| :|:::    ::: |: :: : :    | :::
Db     782 SBSSTTSTTTSSSSTTSTTSSSSTCSTSSCSSSSSSTTSSSSASASSTSTSSTTGCSBSS 723

Qy     483 CGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGG 542
           ::: :  : |::: :||  : ::::  :::   ::  : ::::::::|:|::     ::
Db     722 SBSTSGTSSTSBBTBTTBTSSSSBBSRSSSSSTSSTTBTSBSSSSSBSTSTSSSTTSTSS 663

Qy     543 CGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTA 602
           :::::|:: :::::||     :::: :: ::|::|:  :::::::|: : |::::::|
Db     662 SSSSBTSSTSSBSSTCTTSTTTSSSSCSSSSSTSSTSSSBSSSSSCBTSSCSSSSSSCCT 603

Qy     603 CAACGCAGCCCGC 615
           |  |:|  ||| |
Db     602 CGCCSCSCCCCCC 590


RESULT   7
CNS0181E/c
LOCUS       CNS0181E     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37H05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108764
VERSION     AL108764.1  GI:5629068
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) :
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Bernaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN37H05"
                     /note="end : SP6"
BASE COUNT       61 a    215 c    301 g    235 t    289 others
ORIGIN


  Query Match             3.8%;  Score 46.6;  DB 83;  Length 1101;
  Best Local Similarity   40.2%;  Pred. No. 7.2;
  Matches  101;  Conservative   40;  Mismatches  109;  Indels    1;  Gaps    1;

Qy     389 CCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGC 448
           ||  : ::::  ::  :||   :|: ::  || : |::||| ||:| :||  :||: :
Db     308 CCCCMCVVVMVVGMMAASCCCCAVASCSSCCCCSSCCSSCCGCCGSCCSACSCSCCSCSG 249

Qy     449 TTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGA 508
             |:| :   : ||  |   || |     | |:|| |:|  :   ||:|  :| :: :
Db     248 AAGMCGSMSGVGGCACAAGGGCGGAGACCCGCSCCACSCCCSSSCGCSCSSCSCCSSSSS 189

Qy     509 TGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCG 568
           ||| |||   ||||||| :|  |  |     | |:|   :  | ||  |  :::|| :|
Db     188 CGCGACCCCSSAGGGGCCCSGCCGCCCACCCCGCCSGACCVGACCGCCAC-CMMVGGGSG 130

Qy     569 ACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCA 628
           ||| :  || ||  ||  | |   || |  | ||  | | |||||  |||   ||||
Db     129 ACGCMAAGGGGAGAGGGCCGGNAGCCCCGCNCCACCCCCCCGCCCGCCACGCGCCGGGCA 70

Qy     629 TGGGCGCGACC 639
           | || | |||
Db      69 AGCGCCCAACC 59


RESULT   8
CNS006XK
LOCUS       CNS006XK     935 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4[illegible]5019
KEYWORDS    GSS.
SOURCE      fruit fl[illegible].
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library,
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N09"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
```

```
ORIGIN

  Query Match              3.8%;  Score 46.4;  DB 82;  Length 935;
  Best Local Similarity   27.7%;  Pred. No. 7.7;
  Matches   77;  Conservative   76;  Mismatches  125;  Indels    0;  Gaps    0;

Qy    368 GCACCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCG 427
          :| :|:  : :|   |  :: ::::: |: :|||  : ::|| |::  |   |  |  ||
Db    511 SCSMCRTSSGSTGCSTCCMYMSSSVSCCSCSGTCCGYCSSCTSCKMCSCTYGCKCGCGCG 570

Qy    428 CCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCC 487
          |:  |::::| :: :::  :||:   |::  : |:|  | ||  |: || | :||   |
Db    571 CSTSCSSSSCCSBBSYSTCSCTBCTKCSSGCSTGSGCTGCCGGGGSGCGGCGSGCGCGCG 630

Qy    488 AGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGG 547
          ||  ||  :    :  : :  ::|  | :  :|:||  : || | |  |  | |: :|
Db    631 SGGGCGGGCSSSGGSSGSSSGGSSGSSCGSSSSGSGGGGSGGGCGSGGCGGCGCGSCSGC 690

Qy    548 TGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACG 607
          :| |::   :|  :  |: | || || :  :  ||:  ::|:: |: |:     :  ::
Db    691 GSCGCSSGCGSGCGSGCGSGGGCGSCGCSCSSCGCGSGCSSGSSGCSCCSCCGSCSCSSS 750

Qy    608 CAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACG 643
          : :||:    ||::: :|    : |  |: |||   ||
Db    751 SGSCCSCGCGCGMSMCSGCGGSSCGGCCSGCCGCSSCG 788


RESULT  9
CNS012I3/c
LOCUS       CNS012I3      645 bp    DNA              GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL101589
VERSION     AL101589.1  GI:5613200
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN08C07"
                     /note="end : T7"
BASE COUNT       28 a     26 c     85 g     92 t    414 others
ORIGIN


  Query Match              3.7%;  Score 46.2;  DB 82;  Length 645;
  Best Local Similarity   10.3%;  Pred. No. 8;
  Matches   43;  Conservative  135;  Mismatches  238;  Indels    3;  Gaps    1;

Qy    311 GCCTGCGACACGGGCAGATCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCA 370
          :::  :    |::   | |      ||    || ||   |||::: ::: :    ::  ::
Db    633 SSSNSNNNNASSNNNANANANNNTTTTTTTTNTTTTTTTTSSSNSSSSSCAAASSSNSSN 574

Qy    371 CCGATGCGTTGTTGGATTCCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCG 430
          | |   :   :   :    |:::|| :: :|  : | :     :  :    : :::
Db    573 CAGSNNSCNGSNNNSSNSNCSSVACAGSSSSANCSSACSCSNSNNNSNNNSSNNNSNSSS 514

Qy    431 ACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGG 490
          |     |  |  :   |:     :  :  ::|| ::   : :    :::: |    :
Db    513 ATNNNNCTANAANASSANNTSAANGSSNSNNSSAAANSSNAASASNTTNSSSSNCNNNNS 454

Qy    491 TTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGC 550
          :: :  | |:: :  : :  :: |:   :::: :   : :   :: :  : :  :
Db    453 NNNSSTSSNAAAASSSSCASNSSSSSASASNAASSSSAASSANSNSSNSASSASANSSAS 394

Qy    551 CCGGCGTCGAACCGGCCGACGTCGTGGTGATCGG---CGCCGGCACCGCCGGCTACAACG 607
          :::::: ::  :|  :: | : :  :   :   :        ::: ::::::::| : | |:
Db    393 SSSSSSSSSCSSCNASSAATSNSNNSNSAVSNNSTNNNNNNSSSNSSSSSSSCSVAACCS 334

Qy    608 CAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACA 667
          :  ::::|||:  : |: :    : ::::  :     ::     :     |      :|
Db    333 SSNSSSSCATSNNSSASNSANAASASSSSNNSNNSNASSNNNNSSNNNSNANNSNNNSSA 274

Qy    668 AACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCC 726
            :::|    :   :::       :::|:::             :     |   : :
Db    273 NSNSSSSSCSCNNSNSNNSSSCCSNNNNSSSCSSSNNNNSNNNSNNSNNNNNANNNSNS 215


RESULT  10
CNS006XK/c
LOCUS       CNS006XK      935 bp    DNA              GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neopte  ; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephy  oidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N09"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
ORIGIN
```

```
  Query Match             3.7%;  Score 45.8;  DB 82;  Length 935;
  Best Local Similarity   27.7%;  Pred. No. 10;
  Matches   94;  Conservative   83;  Mismatches  162;  Indels    0;  Gaps    0;

Qy     407 CCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAG 466
           |:  || : ::: |    : |:|| | |:   :  ||::  |  ||| :   | ||
Db     856 CSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGSCSCCSSSCCGCGCCGSCSCGCGCCGGC 797

Qy     467 TCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCC 526
             |||:| || : || |:  | |:   ||: :  |   | :  .|:  : :   ||:| :
Db     796 SGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKSKCGCGCGSGGSCSSSSGSGSCGGSGGS 737

Qy     527 GCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCG 586
           ||:: : ||    |:: :| |   || :|| :   :|| ::|  :    |: |  | ||
Db     736 GCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSCGCSCGCSSGCGSCGCSGSCGCGCCGCC 677

Qy     587 CCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGG 646
           :|| | :| ||: :  :  ::| :||::  : |:     |    :|||| |:   ||
Db     676 SCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSCCSSSGCCCGCCCSCGCGCGCSCGCCGC 617

Qy     647 TTCTAGACATCAACATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCC 706
             | | || |     |:  :| :  :| :: : :| ::::|    |    |:  :| :
Db     616 SCCCCGGCAGCSCASGCSSGMAGVAGSGASRSVVSGGSSSSGSASGCGCGCGMGCRAGSG 557

Qy     707 ACACTCGCTACTCATCGGCCTACGAGCTCGAGGGTGCCG 745
           : :  :: :|  | :|:    :: : : ||:|  :|::
Db     556 KMGSAGSSGRCGGACSGSGGSBSSSKRKGGASGCASCSS 518
```

```
RESULT  11
AW128786
LOCUS       AW128786      591 bp    mRNA            EST       25-OCT-1999
DEFINITION  fe37d05.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
            TR:Q16796 Q16796 NAD(P) TRANSHYDROGENASE PRECURSOR ;, mRNA
            sequence.
ACCESSION   AW128786
VERSION     AW128786.1  GI:6116693
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5422375.
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Wash ngton University Ge: ome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']};
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      119 a    169 c    156 g    147 t
ORIGIN
```

```
  Query Match             3.7%;  Score 45.4;  DB 69;  Length 591;
  Best Local Similarity   50.2%;  Pred. No. 11;
  Matches  112;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy       2 TCTTGCAGATTAATCGAACTTTCTTCACACTGAAGCGTACAGTATCGAGAGGGGTAATCA 61
           ||||   ||   ||| | |  ||  |   || |   || |||           |
Db     185 TCTTTAGGACACATCAAGCCCTCAACCGCCTCACTAGTCCAGGCATTCCTTACAAACAGC 244

Qy      62 TGCGCGTCGGTATTCCGACCGAGACCAAAAACAACGAATTCCGGGTGGCCATCACCCCGG 121
           ||  ||||||  |||| |  || | |    | || ||    || |||||||| | || |
Db     245 TGACCGTCGGCGTTCCTAAGGAAATCTTTCAGAATGAGCGTCGAGTGGCCATCTCTCCTG 304

Qy     122 CCGGCGTCGCGGAACTAACCCGTCGTGGCCATGAGGTGCTCATCCAGGCAGGTGCCGGAG 181
           |||| || |  |  || | |   |  |||  | | ||| |  |  || | || || ||||
Db     305 CCGGAGTGGAAGCTCTCATCAAGCAGGGCTTTAATGTGGTGGTGGAGTCCGGCGCGGGAG 364

Qy     182 AGGGCTCGGCTATCACCGACGCGGATTTCAAGGCGGCAGGCGC 224
           |    |     || | || |     | ||     ||||| ||
Db     365 AATCTGCCAAGTTCTCTGATGACATGTACACAAAAGCAGGAGC 407
```

```
RESULT  12
CNS017ZK
LOCUS       CNS017ZK      744 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37B10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108698
VERSION     AL108698.1  GI:5629002
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
```

d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES Location/Qualifiers
source 1. .744
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37B10"
/note="end : SP6"

BASE COUNT 20 a 191 c 245 g 105 t 183 others
ORIGIN

Query Match 3.7%; Score 45.2; DB 83; Length 744;
Best Local Similarity 34.0%; Pred. No. 13;
Matches 87; Conservative 41; Mismatches 128; Indels 0; Gaps 0;

Qy 329 TCTTGTTCACGTTCTTGCATTTGGCCGCGTCACGTGCTTGCACCGATGCGTTGTTGGATT 388
Db 191 YCYTTGTCGCYTYSGGGCYGSGCCYCYCGYCCSYYBYYTYTCCCSYGCCTCNYSCCGYTT 250

Qy 389 CCGGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGC 448
Db 251 YCCTYCYCYYYYTGYTCCCCCSCGCGCCYTCCCGGCCCCCYCCCSCCCCCCCCCCCCCCC 310

Qy 449 TTGCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGA 508
Db 311 CGCCCCCCCCCCCGCCCCCCCCCCCNCCCCCCSCCSCCCCNNGCNNNCCNNCCCGCCNNG 370

Qy 509 TGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCG 568
Db 371 CSCGNGSCNGSSGGCCSSSSSSGSGSSSSGGGGGGGSCGGSCSNGSCGCGGGNGNGGGSG 430

Qy 569 ACGTCGTGGTGATCGG 584
Db 431 NGGGNGSGSGTKTKGG 446

RESULT 13
CNS017SY
LOCUS CNS017SY 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108460
VERSION AL108460.1 GI:5628764
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37L08"
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN

Query Match 3.7%; Score 45.2; DB 83; Length 1101;
Best Local Similarity 12.6%; Pred. No. 13;
Matches 40; Conservative 157; Mismatches 121; Indels 0; Gaps 0;

Qy 420 CCAGACCGCCGACGGCGCACTACCCCTGCTTGCCCCGATGAGCGAAGTCGCCGGTCGACT 479
Db 755 CGASGAGGGSGRGGGSGCCASSGVCGAACSSASCSSASMCGVSSGSSCSASCGSCCGVSS 814

Qy 480 CGCCGCCCAGGTTGGCGCTTACCACCTGATGCGAACCCAAGGGGGCCGCGGTGTGCTGAT 539
Db 815 CSAVSASSASSVMSKVASAVASCSAVASGMSAGAVSSSCRSSVASSVSAAASVSSSVSSS 874

Qy 540 GGGCGGGGTGCCCGGCGTCGAACCGGCCGACGTCGTGGTGATCGGCGCCGGCACCGCCGG 599
Db 875 SSSSVVSAAAVSASSSSASSASMAVAAAAAVAVSSVSVASVVSSSSSCSSSSSASVVVSA 934

Qy 600 CTACAACGCAGCCCGCATCGCCAACGGCATGGGCGCGACCGTTACGGTTCTAGACATCAA 659
Db 935 SVAASASSVSSSSSSSVSTSSSASVSSVSAVSMSAVVSSSASSSSSVSVSSVVAVAASAA 994

Qy 660 CATCGACAAACTTCGGCAACTCGACGCCGAGTTCTGCGGCCGGATCCACACTCGCTACTC 719
Db 995 AAAAAAAAAASSSSASAVAVVSASSSSSSASSSSSSSASSSVSSVSSSSCSVVSVVSSSS 1054

Qy 720 ATCGGCCTACGAGCTCGA 737
Db 1055 VSVAVASASASASVSSVA 1072

RESULT 14
AI712257
LOCUS AI712257 315 bp mRNA EST 02-FEB-2000
DEFINITION 605069D09.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION AI712257
VERSION AI712257.1 GI:5006195
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE 1 (bases 1 to 315)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246668.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605069 row: D column: 09.
FEATURES Location/Qualifiers
source 1. .315
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"

/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI; Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"

BASE COUNT 61 a 105 c 106 g 42 t 1 others
ORIGIN

Query Match 3.6%; Score 45; DB 51; Length 315;
Best Local Similarity 50.7%; Pred. No. 12;
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy    391 GGCACCACGTCAATTGCCTACGAGACCGTCCAGACCGCCGACGGCGCACTACCCCTGCTT 450
          ||||  ||||    || | || | |  |  |  |||||     |  ||  ||  |
Db     58 GGCATGGCGTCTCCATCCAAAGACAACAGCTGGGTGGCCGAGCAAACTTTAAACCACCCC 117

Qy    451 GCCCCGATGAGCGAAGTCGCCGGTCGACTCGCCGCCCAGGTTGGCGCTTACCACCTGATG 510
          ||||  |  ||| ||| || |||||    ||    |   |  |||| |  | | |    |
Db    118 GCCCTAAGCAGCCAAGGCGGCGGTCACGGCGATAGCGGCGGCGGCGGTCGCGAGCCACCG 177

Qy    511 CGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGGCCGAC 570
          ||  |   | |||||||| | || ||  | | |   | ||    ||||  | |   | |
Db    178 CGTGCGGCAC GGGCCGCCGCGTTCTTCTTGTAGTCCTTCCTCTGGTC GCCGGCGAGGC 237

Qy    571 GTCGTGGTGATCGGCGCCGGCACCGCCGGCTAC 603
          ||||  |    | | |  | |   || | ||| |
Db    238 GTCGGCGGTGGCCGGGCCAGGCGCGTCCGCTCC 270
```

RESULT 15
CNS010RJ
LOCUS CNS010RJ 846 bp DNA GSS ? JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN04N13 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL099337
VERSION AL099337.1 GI:5610948
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 846)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .846
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN04N13"
/note="end : T7"
BASE COUNT 300 a 47 c 40 g 175 t 284 others
ORIGIN

Query Match 3.6%; Score 45; DB 82; Length 846;
Best Local Similarity 16.2%; Pred. No. 14;
Matches 38; Conservative 109; Mismatches 88; Indels 0; Gaps 0;

```
Qy    506 TGATGCGAACCCAAGGGGGCCGCGGTGTGCTGATGGGCGGGGTGCCCGGCGTCGAACCGG 565
          :||  :|:| ::||| |||  |:|: | |  |  ::::|::  |::::::: :    :::
Db    574 KGAGCSGRARSSAAGAGGGGGGSGSGGCGAGGSSSSSSGSSAVGSSSSSVSSSASSASSS 633

Qy    566 CCGACGTCGTGGTGATCGGCGCCGGCACCGCCGGCTACAACGCAGCCCGCATCGCCAACG 625
          ::| ::    :  || :::: :: :: :::: :    :|  :: :::::::|   ::| ::
Db    634 SSGSSSSAAASAAGASSVSSASSAVSSSSSSASAASSSASASVSSSSSVSAVAASSASSS 693

Qy    626 GCATGGGCGCGACCGTTACGGTTCTAGACATCAACATCGACAAACTTCGGCAACTCGACG 685
          ::  :::::: :  :  :: :  | |:::  :  || ::|:|    |::::||: :: :|
Db    694 SSYSVVSSSSAVAAVAVVSASSACSASVSSSVSGCAASSASASSTSTSSSSAASSSSSSG 753

Qy    686 CCGAGTTCTGCGGCCGGATCCACACTCGCTACTCATCGGCCTACGAGCTCGAGGG 740
          ::  |  : :  :  ::    |:|  ::: :  :| ::: : |:: :: :: :::
Db    754 SSASGASSSSAASAASSSSAAASAASSVSAVAVSASSSSASSASSSSSASSSSSS 808
```

Search completed: June 22, 2000, 12:08:09
Job time: 10403 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:49 ; Search time 783.2 Seconds
(without alignments)
-26.084 Million cell updates/sec

Title: US-09-362-485-26
Perfect score: 21
Sequence: 1 AACGAATTCCAATTCCGGGTG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	21	5	A87629	A87629 Sequence 26
2		21	100.0	21	5	A89754	A89754 Sequence 11
3		21	100.0	720	5	A89759	A89759 Sequence 16
4		21	100.0	1245	5	A87605	A87605 Sequence 2
5		21	100.0	1245	5	A89745	A89745 Sequence 2
6		21	100.0	1260	5	A87604	A87604 Sequence 1
7		21	100.0	1260	5	A89744	A89744 Sequence 1
8		21	100.0	2584	1	MTALADH	X63069 M.tuberculo
9	c	17.4	82.9	1223	35	AF068063	AF068063 Supella l
10		16.8	80.0	20628	3	OCTITINR	X64696 O.cuniculus
11		16.8	80.0	182230	53	AC012365	AC012365 Homo sapi
12	c	16.4	78.1	907	1	PDI7272	AJ007272 Prochloro
13	c	16.4	78.1	91136	44	AC019235	AC019235 Homo sapi
14	c	16.4	78.1	95856	44	AC020304	AC020304 Drosophil
15	c	16.4	78.1	119844	56	AC010044	AC010044 Drosophil
16		16.4	78.1	140617	41	AC008058	AC008058 Homo sapi
17		16.4	78.1	158562	42	AC012146	AC012146 Homo sapi
18		16.4	78.1	213732	2	AE001862	AE001862 Deinococc
19		16.2	77.1	76	14	SYNAMBSE	M13718 Synthetic a
20	c	16.2	77.1	249	11	PTU36521	U36521 Pan troglod
21	c	16.2	77.1	250	11	PTU36534	U36534 Pan troglod
22	c	16.2	77.1	254	11	GGU36524	U36524 Gorilla gor
23	c	16.2	77.1	257	11	PTU36522	U36522 Pan troglod
24	c	16.2	77.1	258	11	GGU36527	U36527 Gorilla gor
25	c	16.2	77.1	259	11	PTU36520	U36520 Pan troglod
26	c	16.2	77.1	260	13	G02114	G02114 chimp-CT Ch
27	c	16.2	77.1	261	11	GGU36528	U36528 Gorilla gor
28	c	16.2	77.1	264	11	GGU36523	U36523 Gorilla gor
29	c	16.2	77.1	264	11	GGU36526	U36526 Gorilla gor
30	c	16.2	77.1	266	13	G02115	G02115 Gorilla-CT
31	c	16.2	77.1	276	11	PPU36531	U36531 Pongo pygma
32	c	16.2	77.1	294	11	GGU36529	U36529 Gorilla gor
33	c	16.2	77.1	304	13	G02116	G02116 Orang-CT Or
34	c	16.2	77.1	305	11	PPU36530	U36530 Pongo pygma
35	c	16.2	77.1	306	11	GGU36525	U36525 Gorilla gor
36	c	16.2	77.1	306	11	PPU36532	U36532 Pongo pygma
37	c	16.2	77.1	308	11	PPU36533	U36533 Pongo pygma
38		16.2	77.1	400	13	G01437	G01437 Dm0369 Spra
39		16.2	77.1	2098	1	MVOCTSB	M21315 M.vannielii
40		16.2	77.1	12511	2	AE002129	AE002129 Ureaplasm
41	c	16.2	77.1	12534	1	AB024936	AB024936 Rhodococc
42	c	16.2	77.1	17048	4	D21070	D21070 Rana catesb
43		16.2	77.1	17772	34	CELF22F1	U41549 Caenorhabdi
44	c	16.2	77.1	38019	34	CEF26A3	Z78419 Caenorhabdi
45	c	16.2	77.1	73518	54	AC022553	AC022553 Homo sapi

ALIGNMENTS

```
RESULT   1
A87629
LOCUS       A87629         21 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 26 from Patent WO9836089.
ACCESSION   A87629
VERSION     A87629.1  GI:6736269
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        6 a      5 c      5 g      5 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          |||||||||||||||||||||
Db      1 AACGAATTCCAATTCCGGGTG 21


RESULT   2
A89754
LOCUS       A89754         21 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 11 from Patent WO9832862.
ACCESSION   A89754
VERSION     A89754.1  GI:6738288
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        6 a      5 c      5 g      5 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 2.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          |||||||||||||||||||||
Db      1 AACGAATTCCAATTCCGGGTG 21


RESULT   3
A89759
LOCUS       A89759         720 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 16 from Patent WO9832862.
ACCESSION   A89759
VERSION     A89759.1  GI:6738291
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      133 a    238 c    226 g    123 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          |||||||||||||||||||||
Db     21 AACGAATTCCAATTCCGGGTG 41


RESULT   4
A87605
LOCUS       A87605        1245 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9836089.
ACCESSION   A87605
VERSION     A87605.1  GI:6736245
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      238 a    398 c    387 g    222 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          |||||||||||||||||||||
Db     94 AACGAATTCCAATTCCGGGTG 114


RESULT   5
A89745
LOCUS       A89745        1245 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 2 from Patent WO9832862.
ACCESSION   A89745
VERSION     A89745.1  GI:6738279
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1245
```

```
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       238 a    398 c    387 g    222 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db      94 AACGAATTCCAATTCCGGGTG 114


RESULT   6
A87604
LOCUS       A87604       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9836089.
ACCESSION   A87604
VERSION     A87604.1  GI:6736244
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db     109 AACGAATTCCAATTCCGGGTG 129


RESULT   7
A89744
LOCUS       A89744       1260 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9832862.
ACCESSION   A89744
VERSION     A89744.1  GI:6738278
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       243 a    403 c    389 g    225 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 5;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db     109 AACGAATTCCAATTCCGGGTG 129


RESULT   8
MTALADH
LOCUS       MTALADH      2584 bp    DNA             BCT       01-DEC-1993
DEFINITION  M.tuberculosis gene for L-alanine dehydrogenase.
ACCESSION   X63069 S36765
VERSION     X63069.1  GI:44565
KEYWORDS    alanine dehydrogenase; extracellular; intracellular.
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 2584)
  AUTHORS   Andersen,A.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1991) A.B. Andersen, Statens Seruminstitut,
            Artillerivej 5, DK-2300 Copenhagen S, DENMARK
REFERENCE   2  (bases 1 to 2584)
  AUTHORS   Andersen,A.B., Andersen,P. and Ljungqvist,L.
  TITLE     Structure and function of a 40,000-molecular-weight protein antigen
            of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 60 (6), 2317-2323 (1992)
  MEDLINE   92267644
FEATURES             Location/Qualifiers
     source          1. .2584
                     /organism="Mycobacterium tuberculosis"
                     /strain="Erdman"
                     /isolate="TMC 107, lot 9A-2"
                     /db_xref="taxon:1773"
                     /clone_lib="lambda gt11"
                     /clone="lambda AA67"
     CDS             76. .1197
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="alanine  dehydrogenase"
                     /protein_id="CAA44791.1"
                     /db_xref="GI:44566"
                     /db_xref="SWISS-PROT:P30234"
                     /translation="MRVGIPTETKNNEFQFRVAITPAGVAELTRRGHEVLIQAGAGEG
                     SAITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAAS
                     RACTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGV
                     LMGGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIH
                     TRYSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGC
                     FEGSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAA
                     CRSNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT       463 a    871 c    819 g    431 t
ORIGIN


  Query Match             100.0%;  Score 21;  DB 1;  Length 2584;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db     109 AACGAATTCCAATTCCGGGTG 129


RESULT   9
AF068063/c
LOCUS       AF068063     1223 bp    DNA             INV       11-NOV-1998
DEFINITION  Supella longipalpa allatostatin neuropeptide precursor, gene,
            complete cds.
ACCESSION   AF068063
VERSION     AF068063.1  GI:3859880
KEYWORDS    .
```

```
SOURCE      brown-banded cockroach.
  ORGANISM  Supella longipalpa
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
            Blaberoidea; Blattellidae; Plectopterinae; Supella.
REFERENCE   1  (bases 1 to 1223)
  AUTHORS   Belles,X., Graham,L.A., Bendena,W.G., Ding,Q., Edwards,J.P.,
            Weaver,R.J. and Tobe,S.S.
  TITLE     The molecular evolution of the allatostatin precursor in
            cockroaches
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1223)
  AUTHORS   Bendena,W.G. and Graham,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) Biology, Queen's University, Kingston, ON
            K7L 3N6, Canada
FEATURES             Location/Qualifiers
     source          1. .1223
                     /organism="Supella longipalpa"
                     /db_xref="taxon:83902"
     CDS             93. .1145
                     /note="neuropeptide inhibitor of juvenile hormone
                     biosynthesis"
                     /codon_start=1
                     /product="allatostatin neuropeptide precursor"
                     /protein_id="AAC72894.1"
                     /db_xref="GI:3859881"
                     /translation="MPDSRTCISLQAVLLLALLLQLANSAFGTATAPAGSPEEASSNA
                     GSELLSHLEDSENPELDFVKRLYDFGLGKRAYSYVSEYKRLPVYNFGLGKRSKMYGFG
                     LGKRAGSDSRLYSFGLGKRDYDDYYEEDEDEDQQSSGEDIDDSDAVDLVDKRERLYSF
                     GLGKRARPYSFGLGKRAPSSGVQRLYGFGLGKRGGSLYSFGLGKRGGGRLYAFGLGKR
                     PVNSGRQTGSRFNFGLGKRSEDFDLEEDKRFPQDHRFAFGLGKRKLHPVSIEAVRDEE
                     KDNESESKDVSVQEKKNSTTGERVKRSLSASPYDTSASEEDVDEFARLIRRPFNFGLG
                     KRIPMYDFGIGKRSER"
BASE COUNT      343 a    256 c    311 g    313 t
ORIGIN


  Query Match           82.9%;  Score 17.4;  DB 35;  Length 1223;
  Best Local Similarity  94.7%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGT 20
           ||||| ||||||||||||||
Db     280 ACGAAATCCAATTCCGGGT 262


RESULT  10
OCTITINR
LOCUS       OCTITINR    20628 bp    mRNA            MAM       27-FEB-1996
DEFINITION  O.cuniculus mRNA for titin.
ACCESSION   X64696
VERSION     X64696.1  GI:1208921
KEYWORDS    fibronectin type-III repeat; immunoglobulin-C2; myosin-binding;
            titin gene.
SOURCE      Oryctolagus cuniculus.
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 20628)
  AUTHORS   Labeit,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1992) S. Labeit, EMBL, Meyerhofstr 1, 6900
            Heidelberg, FRG
  REMARK    Revised by [3]
REFERENCE   2  (bases 1 to 20415)
  AUTHORS   Labeit,S., Gautel,M., Lakey,A. and Trinick,J.
  TITLE     Towards a molecular understanding of titin
  JOURNAL   EMBO J. 11 (5), 1711-1716 (1992)
  MEDLINE   92258380
REFERENCE   3  (bases 1 to 20628)
  AUTHORS   Labeit,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB 1996) S. Labeit, EMBL, Meyerhofstr 1, 6900
            Heidelberg, FRG
COMMENT     On Feb 29, 1996 this sequence version replaced gi:1724.
            Related sequences: X17329 and X17330.
FEATURES             Location/Qualifiers
     source          1. .20628
                     /organism="Oryctolagus cuniculus"
                     /strain="CE12"
                     /isolate="CE12.contig"
                     /db_xref="taxon:9986"
                     /clone_lib="PCR on total psoas rabbit cDNA"
     CDS             <1. .>20628
                     /codon_start=3
                     /product="titin"
                     /protein_id="CAA45937.1"
                     /db_xref="GI:1208922"
                     /db_xref="SPTREMBL:Q28733"
                     /translation="EYYFRVCAENKVGVGPTIETKTPILAINPIDRPGEPENLHIADK
                     GKTFVYLKWRRPDYDGGSPNLSYHVERRLKGSTDWERVHKGSIKETHYLVDKCVENQI
                     YEFRVQTKNEGGESDWVKTEEVVVKEDLQKPVLDLKLSGVLTVKAGETIRLEAGVRGK
                     PFPEVVWTKDKDATDFTRSPRAKIDTSADSSKFSLTKAKRSDGGKYVVTATNTAGSFW
                     AYATVNVLDKPGPVRNLKIPDVSSDRCTIRWDPPEDDGGLEIQNYILEKCESKRMVWS
                     TYSATVLTPGTTVTRLIEGNEYIFRVRAENKIGTGPPTESKPVIAKTKYDRPGRPDPP
                     EVTKVSKEEMTVVWSPPEYDGGKSITGYYLEKKEKHSVRWVPVNKSAIPERRLKVQNL
                     IPGHEYQFRVKAENEIGVGEPSLPSRPVVAKDPIEPPGPPINLKVVDTTKSSITLSWG
                     KPVYDGGAPILGYVVEVRPKIADASPDEGWKRCNAAAQLVRTEFTVTSLDENQEYEFR
                     VCAQNQVGIGRPAELKEAIKPKEILEPPEIDLDASMRKLVVVRAGCPIRLFAIVRGRP
                     APKVTWRKVGIDNVVRKGQVDLVDTMAFLVIPNSTRDDSGKYSLTLVNPAGEKAVFVN
                     VRVLDTPGPVSDLKVSDVTKTSCHVSWAPPENDGGSQVTHYIVEKRDAERKTWSIVNP
                     EVKKTSCQVTNLVPGNEYYFRVTAVNEYGPGVPADVPKPVLASDGLSEPDPPRKLEVT
                     EMTKNSATLAWLPPLRDGGAKIDGYIISYREEDQPADRWTEYSVVKDLSLVITGLKEG
                     KKYKFRVAARNAVGVSLPREAEGVYEAKEQLIPPKILMPEQITIKAGKKLRIEAHVYG
                     KPQPICKWKKGEDDVVTSSHLAVHKAESSSILIIKDVTRKDSGYYSLTAENSSGTDTQ
                     KIKVIVMDRPGPPQPPFDISDIDADACSLSWHIPLEDGGSNITNYIVEKCDVSRGDWV
                     TALASVTKTSCRIGKLIPGQEYVFRVRAENRFGISEPLQSPKMLAQFPFGVPSEPKNA
                     RVTKVNKDCIFVAWDRPDSDGGSPITGYLIERKGRNSLLWVKANDTAVRSTEYPCAGL
                     VEGLEYSFRIYALNKAGSSPPSKPTEYVTARTPVDPPGKPEVIDVTKSTVSLIWARPK
                     HDGGSKIIGYFVEACKLPGDKWVRCNTTPHQIPHEEYTVTGLEENAQYQFRAIAKTAV
                     NISQPSELRTPVTIHAENVPPRIDLSVAMKSLLTVKAGTNVCLDATVFGKPMPTVSWK
                     KEGTVLKPAEGIKMAMQRNLCTLELFSVNRKDSGDYTITAENSSGSKSATIKLKVLDR
                     PGPPASVKINKMYSDRAMLSWEPPLEDGGSEITNYIVDKRETSSRNWAQVSANVPITS
                     CSVEKLIEGHEYQFRICAENKYGVGDPVFTEPAIAKNPYDPPGRCDPPVISNVTKDHM
                     TVSWKPPADDGGSPITGYLLEKRETHAVNWTKVNRKPVIERTIKATGLQEGTEYEFRV
                     TAINKAGPGKPSDASKAVYAQDPLYPPGPPAFPKVYDTTRSSVSLSWGKPAYDGGSPI
                     IGYLVEVKRADTDNWVRCNLPQKLQKTRFEVTGLMENTEYQFRVYAVNKVGYSDPSDV
                     PDKHCPKDILIPPEGELDADLRKTLILRAGVTMRLYVPVKGRPPPKITWSKPNVNLRE
                     RIGLDIKSTDFDTFLRCENVNKYDAGKYILTLENSCGKKGYTIVVKVLDTPGPPVNVT
                     VKEISRDSAYITWDPPIVDGGSPIINYVVEKRDAERKSWSTVTTECSKTSFRVSNLEE
                     GKSYFFRVYAENEYGIGDPGETRDAVKASETPGPVVDLKVLTVTKSSCNIGWKKPRSD
                     GGSRITGYVVDFLTEENKWQRVMKSLSLQYSTKDLNEGKQYTFRVSAENENGEGTPSE
                     ITVVAKDDVVAPDLDLKDLPDLCYLAKENSNFRLKIPHQGKPAPSVTWKKGEDPLATD
                     TRVSVESSAVNTTLVVYDCQKSDAGKYTITLKNVAGTKEGTLSIKVGCKPGIPTGPIK
                     FDEVTAEAITLKWGPPKDDGGSEITNYILEKRDSVNNKWVTCASAVQKTTFRVTRLHE
                     GMEYTFRVSAENKYGVGEGLKSEPIVAKHPFDVPDAPPPPNIVDVRHDSVSLTWTDPR
                     KTGGSPITGYHIEFKERNSLLWKRANKTPIRMKDFKVTGLTEGLEYEFRVMAINLAGV
                     GKPSLPSEPVVALDPIDPPGKPEVINVTRNSVTLIWTEPKYDGGHKLTGYIVEKRDLP
                     SKTWMKANHINVPDCAFTVTDLVEGGKYEFRIRAKNTAGAISAPSESTGTIICKDEYE
                     APTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDITQITSTP
                     TSSMLTVKYASRKDAGEYTITATNPFGTKEEHVRVTVLDVPGPPGPIEISNVSAEKAT
                     LTWTPPLEDGGSPIKSYVLEKRETSRLLWTVVAEDIQSCRHVVTKLIQGNEYLFRVSA
                     VNHYGKGEPVQSEPVKMVDRFGPPGPPGKPEVSNVTKNTATVSWKRPTDDGGSEITGY
                     YVERREKKGLRWVRATKTPVSDLRCKVTGLQEGNTYEFRVSAENRAGIGPPSDASNYV
                     LMKDVAYAPGPPSNARVTDTTKKSASLAWGKPHYDGGLEITGYVVEHQKVGDETWVKD
                     TTGPALRITEFVVPDLHTKEKYNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDA
                     ELRRTLVVRAGLSIRIFVPIKGRPAPEVTWTKDDINLKTRANIENTESFTLLIIPECN
                     RYDTGKFVMTIENPAGKKSGFVNVRVLDTPGPVLNLRPTDITKDSVTLHWDLPLIDGG
                     SRITNYIVEKREATRKSYSTVTTKCHKCTYKVTGLSEGCEYFFRVMAENEYGIGEPSE
                     TKEPVKASEAPSPPDSLNIMDITKSTVSLAWPKPKHDGGSKITGYVIEAQRKGSDQWT
                     HITTVKGLECVVRNLTEGEEYTFQVMAVNSAGRSAPRESRPVIVKEQTMLPELDLRGI
                     YQKLVIAKAGDNIKVEIPVLGRPKPTVTWKKGDQVLKQTQRVNVENTATSTILNISEC
                     VRSDSGPYPLTAKNIVGEVGDVITIQVHDIPGPPTGPIKFDEVSSDFVTFSWEPPEND
                     GGVPISNYVIEMRQTDSTTWVELATTVIRTTYKATRLTTGVEYQFRVKAQNRYGVGPG
                     ITSASIVANYPFKVPGPPGTPQVTAVTKDSMTISWHEPLSDGGSPILGYHVERKERNG
```

ILWQTVSKALVPGNIFKSSGLTDGIAYEFRVIAENMAGKSKPSKPSEPVLALDPIDPP
GKPIPLNITRHTVTLKWAKPEYTGGFKITSYIVEKRDLPNGRWLKANFSNILENEFTV
SGLTEDAAYEFRVIAKNAAGAISPPSEPSDAITCRDDVEAPRILVDVRFKDTVILKAG
EAFKLEADVSGRPPPTMEWTKDGKELEGTGKLEIKIADFSTYLINKDSSRRDSGAYIL
TATDPGGFAKHIFNVKVLDRPGPPEGPLAVSEVTSEKCVLSWLPPLDDGGAKIEHYIV
QKRETDRLAWTNVASEVQVTKLKVTKLLKGNEYIFRVMAVNKYGVGEPLESEPVLAVN
PYGPPDPPKNPEVTTITKDSMVVCWGHPDSDGGSEIINYIVERRDKAGQRWVKCNKKT
VTDLRFKVSGLTEGHEYEFRIMAENAAGISAPSRTSPFYKACDAVFKPGPPGNPRVLD
TSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWKIVTPPAGLKATSYTITNLVEN
QEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADLRKLVVIRACCTLRLFV
PIKGRPDPEVKWTREHGESLDKASIESTSSYTLLIVGNVNRFDSGKYILTVENSSGSK
SAFVNVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIKNYIVEKRESTRKAY
STVATNCHKTSWKVDQLQEGSSYYFRVLAENEYGIGLPAETAESVKASERPLPPGKIT
LVDVTRNSVSLSWEKPEHDGGSRILGYIVEMQSKGSDKWATCATVKVTEATITGLIQG
EEYSFRVSAQNEKGISDPRQLSVPVIAKDLVIPPAFKLLFNTFTVLAGEDLKIDVPFI
GRPTPTVTWHKDDVPLKQTTRVNAESTENSSLLSIKEACREDVGHYVVKLSNSAGEAT
ETLNAIILDKPGPPTGPVKMDEVTADSITISWEPPKYDGGSSINNYIVEKRDTSTTTW
QIVSATVARTTIKASRLKTGCEYQFRIAAENRYGKSTYLNSEPIVAQYPFKVPGPPGT
PFVTLSSRDSMEVQWNEPVNDGGSRVIGYHLERKERNSILWVKLNKTPIPQTKFKTTG
LEEGIEYEFRVSAENIVGIGKPSKVSECYVARDPCDPPGRPERIIVTRNSVTLQWKKP
TYDGGSKITGYVVEKKELPDGRWMKASFTNIMDTQFEVTGLVEDHRYEFRVIARNAAG
VFSEPSESTGAITARDEIDPPRISMDPKYKDTIVVHAGESFRIDADIYGKPIPTTQWI
KGDQELSNTARLEIKSTDFATSLSVKDAFRVDSGNYVLKAQNVAGERSVTVNVKVLDR
PGPPEGPIVISGVTAEKCTLAWKPPLQDGGSDIINYIVERRETSRLVWTVVDANVQTL
SCKVTKLLEGNEYIFRVMAVNKYGVGEPLESEPVIAKNPFVVPDAPKAPECTTVTKDS
MIVVWERPASDGGSEILGYVLEKRDKEGIRWTRCHKRLIGELRLRVTGLIENHNYEFR
VSAENAAGLSEPSPPSAYQKACDPIYKPGPPNNPKVMDITRSSVFLSWSKPIYDGGCE
IQGYIVEKCDVSVGEWTMCTPPTGINKTNIEVEKLLEKHEYNFRICAVNKAGVGDHAD
VPGPVIVEEKLEAPDIDLDLELRKIINIRAGGSLRLFVPIKGRPTPEVKWGKVDGEIR
DAAIIDSTSSFTSLVLDNVNRYDSGKYTLTLENSSGTKSAFVTVRVLDTPSPPVNLKV
TEITKDSVSITWEPPLLDGGSKIKNYIVEKRDSTRKSYAAVVTNCHKSSWKIDQLQEG
CSYYFRVTAENEYGIGLPARTADPIKVAEVPQPPGKITVDDVTRNSVSLSWTKPEHDG
GSKIIQYIVEMQAKHSEKWSECARVKSLEAVITNLTQGEEYLFRVVAVNEKGRSDPRS
LAVPIVAKDLVIEPDVKPAFSSYSVQVGQDLKIEVPISGRPKPTITWTKDGLPLKQTT
RINVADSLDLTTLSIKETHKDDSGHYGITVANVVGQKTASIEIITLDKPDPPKGPVKF
DEVSAESITLSWNPPLYTGGCQITNYIV KRDTTTTVWDVVSATVARTTLKVTKLKTG
TEYQFRIFPENRYGQSFALDSEPIVAQYPYKEPGPPGTPFVTATSKDSMVVQWHEPIN
NGGSPIIGYHLERKERNSILWTKVDKSIIHDTQFKALNLEEGIEYEFRVYAENIVGVG
KASKNSECYVARDPCDPPGTPEAIIVKRNEITLQWTKPVYDGGSMITGYIVEKRDLPE
GRWMKASFTNVIETQFTVSGLTEDQRYEFRVIAKNAAGTMSKPSDSTGPITAKDEVEL
PRISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIEWLRGDKEVEESARCEIKNTDFK
ALLIVKDAIRIDGGQYILRASNVAGSKSFPVNVKVLDRPGPPEGPVQVTGVTCEKCTL
TWSPPLQDGGSDIPHYVVEKRETSRLAWTVVASEVVTNSLKITKLLEGNEYIFRIMAV
NKYGVGEPLESAPVLMKNPFVVPGPPKSLEVTNIAKDSMTVCWNRPDSGGGSEITGYI
VEKRDRSGIRWIKCNKRRVTDLRFRVTGLTEDHEYEFRVSAENAAGVGEPSPATVYYK
ACDPVFKPGPPTNAHVVDTTKNSITLAWGKPIYDGGSEVLGYIIEICKADEEEWQIVT
PQTGLKANRFEISKLIEHQEYKIRVCALNKVGLGEAASVPGTVKPEDKLEAPELDLDS
ELRKGIVVRAGGSARIHIPFKGRPTPDITWSREEGEFTDKVQVEKGVNFTQLSIDNCD
RNDAGKYIVKLENSSGTKTAFVTVKVLDTPGPPQNLAVKEVKKDSAVLVWEPPIIDGG
AKVRNYVIDKRESTRKAYANVSSKCNKTTFKVENLTEGAIYYFRVMAENEFGVGVPVE
TVDAVKAAEPPSPPGKVTLTDVSQTSASLMWEKPEHDGGSRVLGYVVEMQPKGTEKWS
VVAESKVCNAVVTGLSSGHEYQFRVKAYNEKGKSDPRVLGVPVIAKDLTIQPSFKLPF
KRYSVQAGEDLKIEIPVIGRPRPEIFWVKDGEPLRQTTRVNVEETATSTILHIKESSK
DDFGKYTITATNSAGTATENLSVIVLEKPGPPVGPVRFDEISADFVVLSWEPPAYTGG
CQISNYIVEKRDTTTTWHIVSATVARTTIKVTKLKTGSEYQFRIYAENRYGKSTSLD
SKPVIVQYPPFKEPGPPGTPFVTSVSRDQMLVQWHEPVNDGGSKVLGYHLEQKEKNSIL
WVKVNKTLIQDTKFKTTGLDEGLEYEFKVSAENIVGIGKPSKVSECFVARDPCDPPGR
PEAIVITRNNVTLKWKKPAYDGGSKITGYIVEKKDLPDGRWMKASFTNVLETEFTVSG
LVEDQRYEFRVIARNAAGNLSEPSESSGAITARDEIDAPNASLDPKYKDVIVVHAGET
FVLEADIRGKPIPDVVWLKDGKELEETTARMEIKSTIQKTTLVVKDCIRTDGGQYVLK
LSNVGGTKSLPITVKVLDRPGPPEEPLKVSGVTAEKCYLAWNPPLQDGGASISHYIIE
KRETSRLSWTQVSTEVQALNYKVTKLLPGNEYIFRVMAVNKYGIGEPLESEPVIACNP
YKPPGPPSTPEASAITKDSMVVTWARPVDDGGAEIEGYILEKRDK"

BASE COUNT 6439 a 4481 c 4950 g 4758 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 3; Length 20628;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ACGAATTCCAATTCCGGGTG 21
||||||| ||| ||||||||||

Db 18634 ACGAATACCAGTTCCGGGTG 18653

RESULT 11
AC012365
LOCUS AC012365 182230 bp DNA HTG 01-FEB-2000
DEFINITION Homo sapiens clone RP11-459C22, WORKING DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION AC012365
VERSION AC012365.2 GI:6806870
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182230)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182230)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jan 28, 2000 this sequence version replaced gi:6114946.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0459C22
---------- Summary Statistics ----------
Sequencing vector: M13; 83%
Sequencing vector: plasmid; 17%
Chemistry: Dye-primer ET; 83% of reads
Chemistry: Dye-terminator Big Dye; 17% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174226 bases at least Q40
Consensus quality: 177429 bases at least Q30
Consensus quality: 179357 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 182230; sum-of-contigs
Quality coverage: 3.77 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1568: contig of 1568 bp in length
* gap of unknown length
* 1569 3821: contig of 2253 bp in length
* gap of unknown length
* 3822 6003: contig of 2182 bp in length
* gap of unknown length
* 6004 8691: contig of 2688 bp in length
* gap of unknown length
* 8692 11369: contig of 2678 bp in length
* gap of unknown length
* 11370 13199: contig of 1830 bp in length
* gap of unknown length
* 13200 16033: contig of 2834 bp in length
* gap of unknown length
* 16034 18134: contig of 2101 bp in length
* gap of unknown length
* 18135 21702: contig of 3568 bp in length
* gap of unknown length

```
            *    21703    25495: contig of 3793 bp in length
            *                    gap of unknown length
            *    25496    28377: contig of 2882 bp in length
            *                    gap of unknown length
            *    28378    36037: contig of 7660 bp in length
            *                    gap of unknown length
            *    36038    43347: contig of 7310 bp in length
            *                    gap of unknown length
            *    43348    50768: contig of 7421 bp in length
            *                    gap of unknown length
            *    50769    59163: contig of 8395 bp in length
            *                    gap of unknown length
            *    59164    69128: contig of 9965 bp in length
            *                    gap of unknown length
            *    69129    81202: contig of 12074 bp in length
            *                    gap of unknown length
            *    81203    92419: contig of 11217 bp in length
            *                    gap of unknown length
            *    92420   106102: contig of 13683 bp in length
            *                    gap of unknown length
            *   106103   119345: contig of 13243 bp in length
            *                    gap of unknown length
            *   119346   134231: contig of 14886 bp in length
            *                    gap of unknown length
            *   134232   154901: contig of 20670 bp in length
            *                    gap of unknown length
            *   154902   182230: contig of 27329 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182230
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-459C22"
BASE COUNT    57545 a   34623 c   34339 g   55712 t     11 others
ORIGIN

  Query Match             80.0%;  Score 16.8;  DB 53;  Length 182230;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGT 20
           ||| |||||||||| ||||
Db   12746 AACAAATTCCAATTCAGGGT 12765


RESULT  12
PDI7272/c
LOCUS       PDI7272       907 bp    DNA             BCT       03-MAY-1999
DEFINITION  Prochloron didemni gap3 gene, partial.
ACCESSION   AJ007272
VERSION     AJ007272.1  GI:4741692
KEYWORDS    gap3 gene; glyceraldehyde-3-phosphate dehydrogenase.
SOURCE      Prochloron didemni.
  ORGANISM  Prochloron didemni
            Bacteria; Cyanobacteria; Prochlorophytes; Prochloraceae;
            Prochloron.
REFERENCE   1  (bases 1 to 907)
  AUTHORS   Figge,R.M., Schubert,M., Brinkmann,H. and Cerff,R.
  TITLE     Glyceraldehyde-3-phosphate dehydrogenase gene diversity in
            eubacte ia and eukaryotes: evidence 'or intra- and inter-kingdom
            gene transfer
  JOURNAL   Mol. Biol. Evol. 16 (4), 429-440 (1999)
  MEDLINE   99261655
REFERENCE   2  (bases 1 to 907)
  AUTHORS   Figge,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Figge R.M., Institute for Genetics,
            TU-Braunschweig, Spielmannstr. 7, Braunschweig, 38106, GERMANY
FEATURES             Location/Qualifiers
     source          1. .907
                     /organism="Prochloron didemni"
                     /db_xref="taxon:1216"
     gene            1. .907
                     /gene="gap3"
     CDS             <1. .>907
                     /gene="gap3"
                     /codon_start=2
                     /transl_table=11
                     /product="glyceraldehyde-3-phosphate dehydrogenase"
                     /protein_id="CAB41846.1"
                     /db_xref="GI:4741693"
                     /translation="RMGRLALRAGWEFPELEFVHINEIQGEAVTAAHLLEFDSVHGRW
                     DKSVVGEGDRILIEGKSLSFSQYSTPGEVPWDELGVDLVLESSGEFRTPQTLEPYFKR
                     GVSKVIVAAPVKEEALNIVMGVNDHLYEPEKHHLLTAASCTTNCLAPIVKVIHENLGI
                     RHGVITTIHDVTNTQTIVDKPHKDLRRARSCLQSLIPTTTGSATAIALIYPELKGKLN
                     GVAVRVPLLNASLTDCVFEVERKTTVAEVNSFLQGATEKEPLAGILGYEERPLVSIDY
                     KNDPRSSIIDALSTMVIDDTQVKILA"
BASE COUNT      260 a    188 c    207 g    252 t
ORIGIN

  Query Match             78.1%;  Score 16.4;  DB 1;  Length 907;
  Best Local Similarity   94.4%;  Pred. No. 4.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGG 19
           || ||||||||||||||||
Db      57 ACAAATTCCAATTCCGGG 40


RESULT  13
AC019235/c
LOCUS       AC019235    91136 bp    DNA             HTG       30-DEC-1999
DEFINITION  Homo sapiens clone RP11-665C14, *** SEQUENCING IN PROGRESS ***, 31
            unordered pieces.
ACCESSION   AC01 235
VERSION     AC01 235.1  GI:6648280
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 91136)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 91136)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     Center project name: H_NH0565C14.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2445: contig of 2445 bp in length
            *     2446     2465: gap of unknown length
            *     2466     4837: contig of 2372 bp in length
            *     4838     4857: gap of unknown length
            *     4858     8857: contig of 4000 bp in length
            *     8858     8877: gap of unknown length
            *     8878    11120: contig of 2243 bp in length
            *    11121    11139: gap of unknown length
            *    11140    15139: contig of 4000 bp in length
            *    15140    15158: gap of unknown length
            *    15159    17291: contig of 2133 bp in length
            *    17292    17310: gap of unknown length
            *    17311    21310: contig of 4000 bp in length
            *    21311    21329: gap of unknown length
            *    21330    25329: contig of 4000 bp in length
```

```
            *    25330    25348: gap of unknown length
            *    25349    27627: contig of 2279 bp in length
            *    27628    27646: gap of unknown length
            *    27647    31646: contig of 4000 bp in length
            *    31647    31665: gap of unknown length
            *    31666    33825: contig of 2160 bp in length
            *    33826    33844: gap of unknown length
            *    33845    35893: contig of 2049 bp in length
            *    35894    35912: gap of unknown length
            *    35913    39710: contig of 3798 bp in length
            *    39711    39729: gap of unknown length
            *    39730    43729: contig of 4000 bp in length
            *    43730    43748: gap of unknown length
            *    43749    46194: contig of 2446 bp in length
            *    46195    46213: gap of unknown length
            *    46214    48429: contig of 2216 bp in length
            *    48430    48448: gap of unknown length
            *    48449    52448: contig of 4000 bp in length
            *    52449    52467: gap of unknown length
            *    52468    54508: contig of 2041 bp in length
            *    54509    54527: gap of unknown length
            *    54528    57618: contig of 3091 bp in length
            *    57619    57637: gap of unknown length
            *    57638    59635: contig of 1998 bp in length
            *    59636    59654: gap of unknown length
            *    59655    61916: contig of 2262 bp in length
            *    61917    61935: gap of unknown length
            *    61936    65284: contig of 3349 bp in length
            *    65285    65303: gap of unknown length
            *    65304    68024: contig of 2721 bp in length
            *    68025    68043: gap of unknown length
            *    68044    70578: contig of 2535 bp in length
            *    70579    70597: gap of unknown length
            *    70598    74490: contig of 3893 bp in length
            *    74491    74509: gap of unknown length
            *    74510    77418: contig of 2909 bp in length
            *    77419    77437: gap of unknown length
            *    77438    79916: contig of 2479 bp in length
            *    79917    79935: gap of unknown length
            *    79936    82321: contig of 2386 bp in length
            *    82322    82340: gap of unknown length
            *    82341    85388: contig of 3048 bp in length
            *    85389    85407: gap of unknown length
            *    85408    88989: contig of 3582 bp in length
            *    88990    89008: gap of unknown length
            *    89009    91136: contig of 2128 bp in length.
FEATURES             Location/Qualifiers
     source          1. .91136
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-665C14"
BASE COUNT    28533 a  17045 c  16872 g  28113 t    573 others
ORIGIN


  Query Match             78.1%;  Score 16.4;  DB 44;  Length 91136;
  Best Local Similarity   94.4%;  Pred. No. 3.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAATTCCAATTCCGGGTG 21
           |||||||||| |||||||
Db   59363 GAATTCCAATGCCGGGTG 59346


RESULT  14
AC020304/c
LOCUS       AC020304    95856 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020304
VERSION     AC020304.1  GI:6664593
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 95856)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212707 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .95856
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    27424 a  20658 c  20289 g  27485 t
ORIGIN


  Query Match             78.1%;  Score 16.4;  DB 44;  Length 95856;
  Best Local Similarity   94.4%;  Pred. No. 3.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGG 18
           |||||||||| ||||||
Db   24854 AACGAATTCCACTTCCGG 24837


RESULT  15
AC010044/c
LOCUS       AC010044    119844 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-7A8, *** SEQUENCING IN
            PROGRESS ***, 67 unordered pieces.
ACCESSION   AC010044
VERSION     AC010044.5  GI:6996730
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 119844)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.B., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119844)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

COMMENT On Feb 18, 2000 this sequence version replaced gi:5903024.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: DRCV
Center clone name: RPCI98-7A8
---------- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 27% of reads
Chemistry: Dye-terminator Big Dye: 30% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 57440 bases at least Q40
Consensus quality: 74844 bases at least Q30
Consensus quality: 87451 bases at least Q20
Estimated insert size: 107883; sum-of-contigs estimation
Quality coverage: 0.7x in Q20 bases; sum-of-contigs estimation
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 815: contig of 815 bp in length
* 816 835: gap of unknown length
* 836 1822: contig of 987 bp in length
* 1823 1842: gap of unknown length
* 1843 2692: contig of 850 bp in length
* 2693 2712: gap of unknown length
* 2713 3756: contig of 1044 bp in length
* 3757 3776: gap of unknown length
* 3777 4964: contig of 1188 bp in length
* 4965 4984: gap of unknown length
* 4985 5912: contig of 928 bp in length
* 5913 5932: gap of unknown length
* 5933 6598: contig of 666 bp in length
* 6599 6618: gap of unknown length
* 6619 7639: contig of 1021 bp in length
* 7640 7659: gap of unknown length
* 7660 8578: contig of 919 bp in length
* 8579 8598: gap of unknown length
* 8599 9482: contig of 884 bp in length
* 9483 9502: gap of unknown length
* 9503 10504: contig of 1002 bp in length
* 10505 10524: gap of unknown length
* 10525 11628: contig of 1104 bp in length
* 11629 11648: gap of unknown length
* 11649 12482: contig of 834 bp in length
* 12483 12502: gap of unknown length
* 12503 13523: contig of 1021 bp in length
* 13524 13543: gap of unknown length
* 13544 14709: contig of 1166 bp in length
* 14710 14729: gap of unknown length
* 14730 16426: contig of 1697 bp in length
* 16427 16446: gap of unknown length
* 16447 17689: contig of 1243 bp in length
* 17690 17709: gap of unknown length
* 17710 18903: contig of 1194 bp in length
* 18904 18923: gap of unknown length
* 18924 20328: contig of 1405 bp in length
* 20329 20348: gap of unknown length
* 20349 21841: contig of 1493 bp in length
* 21842 21861: gap of unknown length
* 21862 23815: contig of 1954 bp in length
* 23816 23835: gap of unknown length
* 23836 24833: contig of 998 bp in length
* 24834 24853: gap of unknown length
* 24854 26523: contig of 1670 bp in length
* 26524 26543: gap of unknown length
* 26544 27640: contig of 1097 bp in length
* 27641 27660: gap of unknown length
* 27661 29254: contig of 1594 bp in length
* 29255 29274: gap of unknown length
* 29275 30499: contig of 1225 bp in length
* 30500 30519: gap of unknown length
* 30520 31355: contig of 836 bp in length
* 31356 31375: gap of unknown length
* 31376 32484: contig of 1109 bp in length
* 32485 32504: gap of unknown length
* 32505 33834: contig of 1330 bp in length
* 33835 33854: gap of unknown length
* 33855 35661: contig of 1807 bp in length
* 35662 35681: gap of unknown length
* 35682 36972: contig of 1291 bp in length
* 36973 36992: gap of unknown length
* 36993 38589: contig of 1597 bp in length
* 38590 38609: gap of unknown length
* 38610 40636: contig of 2027 bp in length
* 40637 40656: gap of unknown length
* 40657 41526: contig of 870 bp in length
* 41527 41546: gap of unknown length
* 41547 43090: contig of 1544 bp in length
* 43091 43110: gap of unknown length
* 43111 44341: contig of 1231 bp in length
* 44342 44361: gap of unknown length
* 44362 45717: contig of 1356 bp in length
* 45718 45737: gap of unknown length
* 45738 46736: contig of 999 bp in length
* 46737 46756: gap of unknown length
* 46757 48197: contig of 1441 bp in length
* 48198 48217: gap of unknown length
* 48218 49895: contig of 1678 bp in length
* 49896 49915: gap of unknown length
* 49916 51231: contig of 1316 bp in length
* 51232 51251: gap of unknown length
* 51252 52756: contig of 1305 bp in length
* 52757 52775: gap of unknown length
* 52577 54464: contig of 1888 bp in length
* 54465 54484: gap of unknown length
* 54485 55906: contig of 1422 bp in length
* 55907 55926: gap of unknown length
* 55927 57413: contig of 1487 bp in length
* 57414 57433: gap of unknown length
* 57434 59469: contig of 2036 bp in length
* 59470 59489: gap of unknown length
* 59490 61030: contig of 1541 bp in length
* 61031 61050: gap of unknown length
* 61051 62972: contig of 1922 bp in length
* 62973 62992: gap of unknown length
* 62993 65773: contig of 2781 bp in length
* 65774 65793: gap of unknown length
* 65794 67834: contig of 2041 bp in length
* 67835 67854: gap of unknown length
* 67855 69540: contig of 1686 bp in length
* 69541 69560: gap of unknown length
* 69561 70584: contig of 1024 bp in length
* 70585 70604: gap of unknown length
* 70605 73399: contig of 2795 bp in length
* 73400 73419: gap of unknown length
* 73420 75293: contig of 1874 bp in length
* 75294 75313: gap of unknown length
* 75314 77447: contig of 2134 bp in length
* 77448 77467: gap of unknown length
* 77468 79540: contig of 2073 bp in length
* 79541 79560: gap of unknown length
* 79561 82125: contig of 2565 bp in length
* 82126 82145: gap of unknown length
* 82146 84203: contig of 2058 bp in length
* 84204 84223: gap of unknown length
* 84224 86307: contig of 2084 bp in length
* 86308 86327: gap of unknown length

```
            *    86328    89402: contig of 3075 bp in length
            *    89403    89422: gap of unknown length
            *    89423    92732: contig of 3310 bp in length
            *    92733    92752: gap of unknown length
            *    92753    96993: contig of 4241 bp in length
            *    96994    97013: gap of unknown length
            *    97014    99925: contig of 2912 bp in length
            *    99926    99945: gap of unknown length
            *    99946   103944: contig of 3999 bp in length
            *   103945   103964: gap of unknown length
            *   103965   108031: contig of 4067 bp in length
            *   108032   108051: gap of unknown length
            *   108052   114250: contig of 6199 bp in length
            *   114251   114270: gap of unknown length
            *   114271   119844: contig of 5574 bp in length.

  Query Match             78.1%;  Score 16.4;  DB 56;  Length 119844;
  Best Local Similarity   94.4%;  Pred. No. 3.1e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGG 18
           |||||||||||  ||||||
Db   48499 AACGAATTCCACTTCCGG 48482


Search completed: June 22, 2000, 17:58:59
Job time: 10808 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:02:01 ; Search time 175.2 Seconds
(without alignments)
29.989 Million cell updates/sec

Title: US-09-362-485-26
Perfect score: 21
Sequence: 1 AACGAATTCCAATTCCGGGTG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	1245	1	V49626	Mycobacterium tube
2		21	100.0	1260	1	V49510	Mycobacterium sp.
3		21	100.0	1260	1	V49625	Mycobacterium tube
c 4		15.4	73.3	3265	1	Q24789	DNA encoding Marek
5		15.4	73.3	12537	1	T41705	Lymphocyte specifi
6		15.4	73.3	32768	1	X20515	Polynucleotide seq
c 7		15.2	72.4	382	1	X41037	Human secreted pro
8		15.2	72.4	19142	1	X20580	Polynucleotide seq
c 9		15	71.4	100	1	T30894	Primer 24 for 95 k
10		15	71.4	2681	1	T30870	Engineered 95 kD p
11		14.8	70.5	1858	1	T72175	Alzheimer's diseas
c 12		14.8	70.5	1860	1	T95371	Arabidopsis thalia
c 13		14.8	70.5	110000	1	X20248_00	Borrelia burgdorfe
14		14.6	69.5	361	1	Q60861	Human brain Expres
15		14.6	69.5	1928	1	T95762	Arabidopsis SCAREC
16		14.6	69.5	2092	1	T14529	Human adenylyl cyc
17		14.6	69.5	2207	1	T98580	DNA encoding two S
18		14.6	69.5	3808	1	V42977	Streptococcus pneu
c	**19**	**14.6**	**69.5**	**9179**	**1**	**X13246**	**Enterococcus faeca**
20		14.6	69.5	9402	1	Q41345	Human hepatitis C
21		14.6	69.5	9405	1	Q40426	Full-length Hepati
c 22		14.6	69.5	14672	1	V52244	Streptococcus pneu
23		14.4	68.6	400	1	V75761	Staphylococcus aur
c	**24**	**14.4**	**68.6**	**4237**	**1**	**V61487**	**Human secreted pro**
c 25		14.4	68.6	32768	1	X13037	Enterococcus faeca
c 26		14.4	68.6	110000	1	V30458_4	Continuation (5 of
c 27		14.4	68.6	110000	1	X20248_03	Continuation (4 of
c 28		14.4	68.6	110000	1	X20248_08	Continuation (9 of
c 29		14.4	68.6	110000	1	V30459_4	Continuation (5 of
c 30		14.2	67.6	101	1	T20393	Human gene signatu
31		14.2	67.6	915	1	Q38661	Glyceride lipase c
32		14.2	67.6	975	1	Q04494	Sequence encoding
33		14.2	67.6	997	1	Q38663	Glyceride lipase c
34		14.2	67.6	1065	1	T02472	Cowpox virus T2-eq
c 35		14.2	67.6	1339	1	V80610	Kidney injury asso
c 36		14.2	67.6	1785	1	Q54883	AHAS Gene. Aceto:h
37		14.2	67.6	2038	1	Q38662	Glyceride lipase c
c 38		14.2	67.6	2367	1	V80663	Human DNAX toll-li
39		14.2	67.6	2924	1	Q46089	Sequence of human
40		14.2	67.6	2924	1	Q63261	CD26 cDNA clone. P
c 41		14.2	67.6	2966	1	Q11797	Sequence encoding
42		14.2	67.6	2974	1	Q66201	H. influenzae SB33
43		14.2	67.6	14928	1	Q11707	Immunoglobulin D-r
44		14.2	67.6	110000	1	V30458_2	Continuation (3 of
c 45		14.2	67.6	110000	1	V30458_2	Continuation (3 of

ALIGNMENTS

RESULT 1
V49626
ID V49626 standard; DNA; 1245 BP.
AC V49626;
DT 20-NOV-1998 (first entry)
DE Mycobacterium tuberculosis L-Alanine dehydrogenase 40 KD antigen.
KW ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS Mycobacterium tuberculosis.
PN WO9836089-A2.
PD 20-AUG-1998.
PF 29-JAN-1998; E00483.
PR 29-JAN-1997; EP-101338.
PA (FLOH/) FLOHE L.
PI Flohe L, Hutter B, Kolk A, Singh M;
DR WPI; 98-457123/39.
PT Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT - useful for, e.g. for diagnosis, differentiation of strains,
PT monitoring vaccination and identification of mycobacterial
PT inhibitors
PS Disclosure; Fig 3.19; 55pp; German.
CC THe Mycobacterium tuberculosis L-alanine dehydrogenase (AlaDH) is used in
CC the production of kits for diagnosing tuberculosis (TB) and other
CC mycobacterial infections in humans or animals. Kits are used for direct
CC diagnosis of TB on clinical samples (e.g. body fluids) and can
CC differentiate between pathogenic and non-virulent strains, e.g. for
CC identifying pathogens of the M.tuberculosis (M. t.) complex. The kit may
CC also be used to identify substances that inhibit mycobacteria, for
CC combatting epidemics and for vaccination follow-up. Oligonucleotides
CC derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC also for culture confirmation of isolated strains and for chromosome
CC fingerprinting to detect/differentiate between mycobacteria, and for
CC L-alanine-specific biotransformation reactions. AlaDH is an early
CC antigen, present extracellularly after only a few days of growth, making
CC it an ideal drug target.
SQ Sequence 1245 BP; 238 A; 398 C; 387 G; 222 T;

Query Match 100.0%; Score 21; DB 1; Length 1245;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||||||||||||||||||||
Db 94 AACGAATTCCAATTCCGGGTG 114

RESULT 2
V49510
ID V49510 standard; DNA; 1260 BP.
AC V49510;
DT 20-OCT-1998 (first entry)
DE Mycobacterium sp. AlaDH DNA.
KW Alanine dehydrogenase; AlaDH; ADH; diagnosis; tuberculosis; pathogen;
KW swimmers disease; vaccine; epidemic; infection; identification; ss.
OS Mycobacterium sp.
PN WO9832862-A2.
PD 30-JUL-1998.

```
PF   29-JAN-1998; E00484.
PR   29-JAN-1997; EP-101339.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-427958/36.
PT   Nucleic acid encoding alanine dehydrogenase of Mycobacterium marinum
PT   - used for diagnosis of tuberculosis and other mycobacterial
PT   diseases, also for treatment and prevention, for drug screening and
PT   for bio-transformation
PS   Disclosure; Page 11; 57pp; German.
CC   This sequence encodes an alanine dehydrogenase (ADH) protein isolated
CC   from a Mycobacterium sp. This protein is used to diagnose tuberculosis
CC   and other mycobacterial infections (including 'swimmers' disease', caused
CC   by M. marinum, a fish pathogen) in humans or animals. The protein can
CC   also be used for control of epidemics and for vaccination, to screen for
CC   agents with anti-mycobacterial activity, and in bio-transformations that
CC   are specific for L-alanine. Also mycobacteria can be identified by
CC   analysis of genomic ADH sequences. ADH is an antigen that is secreted
CC   early during infection.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match             100.0%;  Score 21;  DB 1;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 0.084;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db     109 AACGAATTCCAATTCCGGGTG 129


RESULT   3
V49625
ID   V49625 standard; DNA; 1260 BP.
AC   V49625;
DT   20-NOV-1998  (first entry)
DE   Mycobacterium tuberculosis Alanine dehydrogenase.
KW   ss; Alanine dehydrogenase; tuberculosis; TB; mycobacteria; L-alanine.
OS   Mycobacterium tuberculosis.
PN   WO9836089-A2.
PD   20-AUG-1998.
PF   29-JAN-1998; E00483.
PR   29-JAN-1997; EP-101338.
PA   (FLOH/) FLOHE L.
PI   Flohe L, Hutter B, Kolk A, Singh M;
DR   WPI; 98-457123/39.
PT   Diagnosis of, e.g. tuberculosis from alanine dehydrogenase activity
PT   - useful for, e.g. for diagnosis, differentiation of strains,
PT   monitoring vaccination and identification of mycobacterial
PT   inhibitors
PS   Claim 13; Fig 2.3; 55pp; German.
CC   THe Mycobacterium tuberculosis alanine dehydrogenase (AlaDH) is in
CC   the production of kits for diagnosing tuberculosis (TB) and other
CC   mycobacterial infections in humans or animals.  Kits are used for direct
CC   diagnosis of TB on clinical samples (e.g. body fluids) and can
CC   differentiate between pathogenic and non-virulent strains, e.g. for
CC   identifying pathogens of the M.tuberculosis (M. t.) complex.  The kit may
CC   also be used to identify substances that inhibit mycobacteria, for
CC   combatting epidemics and for vaccination follow-up. Oligonucleotides
CC   derived from AlaDH are used similarly in diagnostic hybridisation tests,
CC   also for culture confirmation of isolated strains and for chromosome
CC   fingerprinting to detect/differentiate between mycobacteria, and for
CC   L-alanine-specific biotransformation reactions.  AlaDH is an early
CC   antigen, present extracellularly after only a few days of growth, making
CC   it an ideal drug target.
SQ   Sequence   1260 BP;    243 A;    403 C;    389 G;    225 T;


  Query Match             100.0%;  Score 21;  DB 1;  Length 1260;
  Best Local Similarity   100.0%;  Pred. No. 0.084;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||||||||||||
Db     109 AACGAATTCCAATTCCGGGTG 129


RESULT   4
Q24789/c
ID   Q24789 standard; DNA; 3265 BP.
AC   Q24789;
DT   14-NOV-1992  (first entry)
DE   DNA encoding Marek's disease virus MD20 polypeptide.
KW   Antibodies; vaccine; recombinant; poultry; passive immunotherapy;
KW   diagnostic immunoassay; anti-idiotype; antigen; ss.
OS   Marek's disease virus.
FH   Key             Location/Qualifiers
FT   cds             41. .3265
FT                   /*tag=  a
PN   EP-486106-A.
PD   20-MAY-1992.
PF   13-NOV-1991; 202947.
PR   16-NOV-1990; US-615211.
PA   (ALKU ) AKZO NV.
PI   Morgan RW;
DR   WPI; 92-168713/21.
DR   P-PSDB; R24102.
PT   DNA encoding Marek's disease virus polypeptides MD18 and MD20 -
PT   and antibodies and vaccine useful for the protection of poultry
PT   against MDV infection
PS   Claim 1; Page 18; 31pp; English.
CC   DNA was prepd. for the establishment of a lambda EMBL 3 library by
CC   infecting chicken embryo fibroblasts with a tissue-culture adapted
CC   passage of Marek's disease virus (MDV) strain GA, and incubating until
CC   a 90 percent cytopathic effect had developed.  Total DNA was prepd.
CC   by proteinase K digestion and phenol/chloroform extraction.  DNA
CC   was partially digested with Sau3A and a 20 kb fraction isolated by
CC   electrophoresis on 0.8 percent agarose.  DNA fragments were ligated
CC   with BamHI/EcoRI digested lambda EMBL 3 DNA, packaged in vitro and
CC   plated on E. coli strain LE392.  Screening of this library with the
CC   insert from clone pMD21 contg. the viral DNA resulted in the isolation
CC   of lambda GA09 contg. a 17 kb DNA fragment contg. a region between
CC   the S and I2 fragments of MDV which encoded the antigen MD20.  Vectors
CC   and host cells contg. the MDV MD20 gene and MDV polypeptides can be
CC   used in a vaccine to protect poultry against Marek's disease. Anti-
CC   bodies or antiserum raised by the polypeptides may be used in passive
CC   immunotherapy, diagnostic immunoassays and in the generation of anti-
CC   idiotypic antibodies for use in a test kit for Marek's disease.  The
CC   vaccine may also contain immunogens related to other poultry pathogens,
CC   e.g. infectious bronchitis-virus, Newcastle disease-virus or infectious
CC   bursal disease-virus to produce a multivalent vaccine.
CC   See also Q24788.
SQ   Sequence   3265 BP;    957 A;    614 C;    732 G;    962 T;


  Query Match             73.3%;  Score 15.4;  DB 1;  Length 3265;
  Best Local Similarity   94.1%;  Pred. No. 55;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCG 17
           |||||||||||| ||||
Db    1102 AACGAATTCCAAATCCG 1086


RESULT   5
T41705
ID   T41705 standard; DNA; 12537 BP.
AC   T41705;
DT   20-JAN-1997  (first entry)
DE   Lymphocyte specific interferon regulatory factor gene.
KW   Lymphocyte specific interferon regulatory factor; LSIRF; IRF-3; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   unsure          748
FT                   /*tag=  a
FT                   /note= "sequence uncertainty a  base 748"
```

```
FT   unsure          4159
FT                   /*tag=  b
FT                   /note= "sequence uncertainty at base 4159"
FT   unsure          4513
FT                   /*tag=  c
FT                   /note= "sequence uncertainty at base 4513"
FT   unsure          5277
FT                   /*tag=  d
FT                   /note= "sequence uncertainty at base 5277"
FT   unsure          5310
FT                   /*tag=  e
FT                   /note= "sequence uncertainty at base 5310"
FT   unsure          5885
FT                   /*tag=  f
FT                   /note= "sequence uncertainty at base 5885"
FT   unsure          6425
FT                   /*tag=  g
FT                   /note= "sequence uncertainty at base 6425"
FT   unsure          7413
FT                   /*tag=  h
FT                   /note= "sequence uncertainty at base 6425"
FT   unsure          9812
FT                   /*tag=  i
FT                   /note= "sequence uncertainty at base 9812"
FT   unsure          10357
FT                   /*tag=  j
FT                   /note= "sequence uncertainty at vase 10357"
FT   unsure          10564
FT                   /*tag=  k
FT                   /note= "sequence uncertainty at base 10564"
FT   unsure          11713
FT                   /*tag=  l
FT                   /note= "sequenc euncertainty at base 11713"
PN   WO9632477-A1.
PD   17-OCT-1996.
PF   12-APR-1996; CA0231.
PR   14-APR-1995; US-422733.
PR   03-APR-1996; US-611280.
PA   (AMGE-) AMGEN CANADA INC.
PI   Grossman A,  Matsuyama T,  Richardson CD;
DR   WPI; 96-477128/47.
PT   New genes for murine lymphocyte specific interferon regulatory
PT   factor  - used for modulation of lymphocyte activation and
PT   proliferation
PS   Claim 1; Page 51-58; 92pp; English.
CC   A genomic clone (T41705) contg. exons and introns (not identified
CC   in the specification) codes for novel mouse lymphocyte specific
CC   interferon regulatory factor (LSIRF or ISF-3) (see also R99426), a
CC   protein that binds to interferon-stimulated response elements in
CC   DNA.  To obtain the sequence, a 129/J mouse kidney genomic library
CC   was screened with a probe obtd. by PCR amplification (see also
CC   T41707-8) of a LSIRF cDNA clone (T41701).  2 Clones contained
CC   overlapping sequences that permitted the identification of the
CC   LSIRF gene and 5' flanking sequence (T41704).  By modulating LSIRF
CC   gene expression it is possible to suppress or stimulate lymphocyte
CC   activation.
SQ   Sequence   12537 BP;    3105 A;    2906 C;    3009 G;    3505 T;

  Query Match             73.3%;  Score 15.4;  DB 1;  Length 12537;
  Best Local Similarity   94.1%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCG 17
          |||||||||||||||| |
Db   7719 AACGAATTCCAATTCGG 7735


RESULT   6
X20515
ID   X20515 standard; DNA; 32768 BP.
AC   X20515;
DT   05-MAY-1999  (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum.
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI; 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1; Page 261-279; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
SQ   Sequence   32768 BP;    8253 A;    9783 C;    7257 G;    7433 T;

  Query Match             73.3%;  Score 15.4;  DB 1;  Length 32768;
  Best Local Similarity   94.1%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 AATTCCAATTCCGGGTG 21
          |||||| ||||||||||
Db  30379 AATTCCCATTCCGGGTG 30395


RESULT   7
X41037/c
ID   X41037 standard; cDNA; 382 BP.
AC   X41037;
DT   18-JUN-1999  (first en  y)
DE   Human secreted protein  ' EST SEQ ID NO: 249.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906554-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1238.
PR   01-AUG-1997; US-905134.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153784/13.
DR   P-PSDB; Y12204.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT   muscle, muscle and heart tissue
PS   Claim 1; Page 365; 622pp; English.
CC   X40826 to X41093 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y01602 and
CC   Y11994 to Y12260, respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein. The nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy.
CC   The proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used
```

```
CC   for directing extracellular secretion of a polypeptide or the insertion
CC   of a polypeptide into a membrane, or importing a polypeptide into
CC   a cell.
SQ   Sequence   382 BP;    73 A;    129 C;    89 G;    91 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 382;
  Best Local Similarity   85.0%;  Pred. No. 59;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGTG 21
           ||||||||||  || |||||
Db     159 ACGAATTCCAGGTCAGGGTG 140


RESULT   8
X20580
ID   X20580 standard; DNA; 19142 BP.
AC   X20580;
DT   05-MAY-1999  (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum.
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI; 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1; Page 608-619; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
SQ   Sequence   19142 BP;    4629 A;    5539 C;    4716 G;    4238 T;


  Query Match             72.4%;  Score 15.2;  DB 1;  Length 19142;
  Best Local Similarity   85.0%;  Pred. No. 79;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGTG 21
           ||||||||||  | |||||
Db   15818 ACGAATTCCAACACTGGGTG 15837


RESULT   9
T30894/c
ID   T30894 standard; DNA; 100 BP.
AC   T30894;
DT   24-FEB-1997  (first entry)
DE   Primer 24 for 95 kD protein coding sequence.
KW   Tetrahymena; 80 kD protein; 95 kD protein; telomerase; telomeric repeat;
KW   RNA coupled protein; chromosome; telomere repeat synthesis; antibody;
KW   immortalised cell; cancer; eukaryotic microbe; inhibitor; gene therapy;
KW   fungal infection; therapy; diagnosis; protozoan infection; somatic cell;
KW   p80; p95; polymerase chain reaction; primer; amplify; PCR; ss.
OS   Synthetic.
PN   WO9619580-A2.
PD   27-JUN-1996.
PF   18-DEC-1995; U16531.
PR   19-DEC-1994; US-359125.
PA   (COLD-) COLD SPRING HARBOR LAB.
PI   Autexier C,  Collins K,  Greider C,  Hemish JM,  Kobayashi R;
PI   Yang XH;
DR   WPI; 96-309594/31.
PT   Telomerase protein and related DNA, antibodies, transgenic cells,
PT   etc. - for diagnosis and treatment of cancer and infection by
PT   eukaryotic microbes, also new telomerase inhibitors
PS   Claim 36; Fig 6; 56pp; English.
CC   T30871-T30931 represent amplification primers for the coding sequences
CC   for the Tetrahymena 80 kD and 95 kD proteins (p80 and p95), respectively
CC   (see T30867 and T30868 for wild type coding sequences).  T30871-T30904
CC   represent amplification primers for p95.  The proteins encoded by the
CC   amplified sequences are components of telomerase.  The encoded proteins
CC   act, when coupled to RNA, to add telomeric repeats (of the sequence
CC   TTGGGG) to stabilise chromosomes.  The RNA component provides the
CC   template for the telomere repeat synthesis.  Antibodies against the
CC   encoded proteins, can be used as immunoassay reagents for the detection
CC   of the proteins.  The antibodies can also be used to identify
CC   immortalised cells, or predisposition to immortalisation, particularly
CC   cancer, or to diagnose a disease caused by a eukaryotic microbe.
CC   Inhibitors of the encoded proteins (and the proteins themselves) can be
CC   used for therapy or diagnosis.  The inhibitors can be used to treat
CC   infection by fungi and protozoa.  As somatic cells do not generally
CC   require telomerase, these inhibitors should have little or no toxicity to
CC   the host.  The amplified sequences are used to produce the recombinant
CC   protein, or to isolate similar genes from other organisms, while
CC   transformed cells can be used in gene therapy.
SQ   Sequence   100 BP;    17 A;    25 C;    25 G;    33 T;


  Query Match             71.4%;  Score 15;  DB 1;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTC 15
           |||||||||||||||
Db      84 AACGAATTCCAATTC 70


RESULT   10
T30870
ID   T30870 standard; DNA; 2681 BP.
AC   T30870;
DT   24-FEB-1997  (first entry)
DE   Engineered 95 kD protein coding sequence.
KW   Tetrahymena; 80 kD protein; 95 kD protein; telomerase; telomeric repeat;
KW   RNA coupled protein; chromosome; telomere repeat synthesis; antibody;
KW   immortalised cell; cancer; eukaryotic microbe; inhibitor; gene therapy;
KW   fungal infection; therapy; diagnosis; protozoan infection; somatic cell;
KW   p80; p95; ss.
OS   Synthetic.
PN   WO9619580-A2.
PD   27-JUN-1996.
PF   18-DEC-1995; U16531.
PR   19-DEC-1994; US-359125.
PA   (COLD-) COLD SPRING HARBOR LAB.
PI   Autexier C,  Collins K,  Greider C,  Hemish JM,  Kobayashi R;
PI   Yang XH;
DR   WPI; 96-309594/31.
PT   Telomerase protein and related DNA, antibodies, transgenic cells,
PT   etc. - for diagnosis and treatment of cancer and infection by
PT   eukaryotic microbes, also new telomerase inhibitors
PS   Claim 33; Fig 11; 56pp; English.
CC   T30869 and T30870 represent the coding sequences for engineered versions
CC   of the Tetrahymena 80 kD and 95 kD proteins (p80 and p95), respectively
CC   (see T30867 and T30868 for wild type coding sequences).  These sequences
CC   were engineered to contain the optimum codon usage for expression in
CC   E. coli.  The proteins encoded by these sequences are components of
CC   telomerase.  The encoded proteins act, when coupled to RNA, to add
CC   telomeric repeats (of the sequence TTGGGG) to stabilise chromosomes.  The
CC   RNA component provides the template for the telomere repeat synthesis.
CC   Antibodies against the encoded proteins, can be used as immunoassay
CC   reagents for the detection of the proteins.  The antibodies can also be
CC   used to identify immortalised cells, or predisposition to
CC   immortalisation, particularly cancer, or to diagnose a disease caused by
CC   a eukaryotic microbe.  Inhibitors of the encoded proteins (and the
CC   proteins themselves) can be used for therapy or diagnosis.  The
```

```
CC   inhibitors can be used to treat infection by fungi and protozoa.  As
CC   somatic cells do not generally require telomerase, these inhibitors
CC   should have little or no toxicity to the host.  These sequences are used
CC   to produce the recombinant protein, or to isolate similar genes from
CC   other organisms, while transformed cells can be used in gene therapy.
SQ   Sequence   2681 BP;    836 A;    755 C;    632 G;    458 T;

  Query Match             71.4%;  Score 15;  DB 1;  Length 2681;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTC 15
          |||||||||||||||
Db   1841 AACGAATTCCAATTC 1855


RESULT  11
T72175
ID   T72175 standard; cDNA to mRNA; 1858 BP.
AC   T72175;
DT   25-FEB-1998  (first entry)
DE   Alzheimer's disease DNA sequence from plasmid pGCS2232.
KW   Human; brain; Alzheimer's disease; diagnosis; antibody; expression; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             26. .1258
FT                   /*tag= a
FT                   /note= "Protein encoded by DNA from the human brain which
FT                          has different expression from the normal protein
FT                          when in an Alzheimer's disease patient"
PN   WO9721807-A1.
PD   19-JUN-1997.
PF   12-DEC-1996; J03630.
PR   12-DEC-1995; JP-322745.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PI   Hashida H,  Kuga T,  Nakagawa S,  Sakaki Y,  Zhao N;
DR   WPI; 97-332779/30.
DR   P-PSDB; W21578.
PT   DNA sequences with varied expression in normal and Alzheimer
PT   patients - used for diagnosis and treatment of Alzheimer's disease
PS   Claim 3; Page 53-56; 90pp; Japanese.
CC   The present sequence represents a DNA fragment which is complementary to
CC   mRNA sequences whose level of expression in the brains of Alzheimer's
CC   disease patients varies from normal individuals. The present sequence
CC   represents a specifically claimed DNA fragment in which the expression
CC   is augmented to at least twice that in normal cases. The sequence, and
CC   antibodies to the proteins encoded by the sequence, may be used in the
CC   diagnosis, treatment and investigation of Alzheimer's disease.
SQ   Sequence   1858 BP;    507 A;    457 C;    469 G;    425 T;

  Query Match             70.5%;  Score 14.8;  DB 1;  Length 1858;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 CGAATTCCAATTCCGGGT 20
          ||| |||||| |||||||
Db   1036 CGAGTTCCAAGTCCGGGT 1053


RESULT  12
T95371/c
ID   T95371 standard; cDNA; 1860 BP.
AC   T95371;
DT   14-APR-1998  (first entry)
DE   Arabidopsis thaliana epsilon cyclase cDNA.
KW   Epsilon cyclase; E-cyclase; carotenoid; biosynthetic enzyme;
KW   pigment; vector; pATeps; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             109. .1683
FT                   /*tag= a
FT   polyA_signal    1732. .1736
FT                   /*tag= b
PN   WO9736998-A1.
PD   09-OCT-1997.
PF   28-JAN-1997; U00540.
PR   29-MAR-1996; US-624125.
PA   (UYMA-) UNIV MARYLAND BALTIMORE.
PI   Cunningham FX, Sun Z;
DR   WPI; 97-503091/46.
DR   P-PSDB; W35346.
PT   Eukaryotic carotenoid biosynthetic enzymes and related genes -
PT   useful to control ratio of various carotenoid(s) in host and for
PT   production of novel carotenoid pigments
PS   Claim 4; Page 33-35; 89pp; English.
CC   This cDNA sequence codes for Arabidopsis thaliana epsilon cyclase
CC   (see W35346), an enzyme responsible for the formation of epsilon
CC   end-groups in carotenoids.  It was isolated by mass excision and
CC   colour complementation screening of an A. thaliana cDNA library.
CC   Claimed expression vector pATeps comprising the epsilon cyclase
CC   gene is deposited as ATCC 98005.  Claimed isolated DNA sequences
CC   (see T95371-73 and T95389-91) encoding the claimed eukaryotic
CC   carotenoid biosynthetic enzymes epsilon cyclase, beta-carotene
CC   hydroxylase and isopentenyl pyrophosphate isomerase (see
CC   W35346-51) are used in methods for augmenting the accumulation of
CC   carotenoids and for the production of novel and rare carotenoids
CC   in host cells.  Methods are also provided for controlling the ratio
CC   of various carotenoids in a host, and for screening for eukaryotic
CC   genes that encode enzymes of carotenoid biosynthesis and metabolism.
SQ   Sequence   1860 BP;    531 A;    327 C;    448 G;    554 T;

  Query Match             70.5%;  Score 14.8;  DB 1;  Length 1860;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 GAATTCCAATTCCGGGTG 21
          |||||| ||||||| |||
Db   1159 GAATTCGAATTCCGAGTG 1142


RESULT  13
X20248_00/c
WP Sequence split into 10 fragments  LOCUS X20248 Accession X20248
WP     Fragment Name      Begin     End
WP     X20248_00              1     110000
WP     X20248_01         100001     210000
WP     X20248_02         200001     310000
WP     X20248_03         300001     410000
WP     X20248_04         400001     510000
WP     X20248_05         500001     610000
WP     X20248_06         600001     710000
WP     X20248_07         700001     810000
WP     X20248_08         800001     910000
WP     X20248_09         900001     910715
ID   X20248 standard; DNA; 910715 BP.
AC   X20248;
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #1.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858943-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; U12764.
PR   03-SEP-1997; US-057483.
PR   20-JUN-1997; US-050359.
PR   22-JUL-1997; US-053344.
PR   22-JUL-1997; US-053377.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI   White OR;
```

```
DR   WPI; 99-081217/07.
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 157-671; 1128pp; English.
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence   910715 BP;    327171 A;    129646 C;    130753 G;    323091 T;


  Query Match             70.5%;  Score 14.8;  DB 1;  Length 110000;
  Best Local Similarity   88.9%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CGAATTCCAATTCCGGGT 20
           ||||||| |||||| |||
Db   99177 CGAATTCGAATTCCTGGT 99160


RESULT  14
Q60861
ID   Q60861 standard; DNA; 361 BP.
AC   Q60861;
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST00958.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 402; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00958 has an "excellent" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   361 BP;     63 A;     111 C;     119 G;     64 T;


  Query Match             69.5%;  Score 14.6;  DB 1;  Length 361;
  Best Local Similarity   81.0%;  Pred. No. 1.2e+02;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           ||||||| ||| ||| | |||
Db     240 AACGAATACCAGTTCAGCGTG 260


RESULT  15
T95762
ID   T95762 standard; cDNA; 1928 BP.
AC   T95762;
DT   11-MAY-1998  (first entry)
DE   Arabidopsis SCARECROW clone SRPa7.
KW   SCARECROW; SCR; SRPa7 gene; transgenic plant; root; gravitropism;
KW   crop improvement; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   CDS             152. .1723
FT                   /*tag= a
PN   WO9741152-A1.
PD   06-NOV-1997.
PF   25-APR-1997; U07022.
PR   24-APR-1997; US-842445.
PR   26-APR-1996; US-638617.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE,
PI   Pysh L, Wysocka-Diller J;
DR   WPI; 97-549683/50.
DR   P-PSDB; W38187.
PT   DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
PT   cell division and therefore alter root development, or alter plant
PT   stem or hypocotyl gravitropism
PS   Claim 6; Page 127-128; 221pp; English.
CC   This sequence comprises a partial cDNA clone of the Arabidopsis
CC   SCARECROW (SCR) SRPa7 gene, identified by homology to the
CC   Arabidopsis SCR gene (see T95753).  It encodes a 523-amino acid
CC   polypeptide (see W38187).  SCR is expressed specifically in embryo
CC   root progenitor tissue and in certain root and stem tissues.  It
CC   regulates a specific asymmetric division, and controls gravitropic
CC   response in aerial structures and root formation.  Nucleic acid
CC   molecules (see T95753-66) encoding SCR protein sequences (see
CC   W38178-201), SCR proteins lacking 1 to 4 of MOTIFs I to VI, or
CC   SCR MOTIF 1, II, III, IV, V or VI are claimed.  Transgenic plants
CC   can be engineered to overexpress SCR, so that cell division is
CC   increased in roots, resulting in thicker root development, while a
CC   plant with an altered stem or hypocotyl gravitropism is less
CC   susceptible to lodging.  Plants that contain an antisense
CC   molecule that suppresses the expression of endogenous SCR gene
CC   product show thinner root development.  A gene of interest can be
CC   placed under control of a SCR promoter and expressed in a plant to
CC   confer herbicide, salt, pathogen or insect resistance, or when
CC   expressed in stems to increase starch, lignin or cellulose
CC   biosynthesis (all claimed).
SQ   Sequence   1928 BP;    520 A;    380 C;    500 G;    528 T;


  Query Match             69.5%;  Score 14.6;  DB 1;  Length 1928;
  Best Local Similarity   81.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           ||| ||||| |||||||  ||
Db     457 AACAAATTCAAATTCCGATTG 477


Search completed: June 22, 2000, 18:02:06
Job time: 10138 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:04:11 ; Search time 123.79 Seconds
(without alignments)
22.051 Million cell updates/sec

Title: US-09-362-485-26
Perfect score: 21
Sequence: 1 AACGAATTCCAATTCCGGGTG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	15.4	73.3	3265	1	US-07-912-015-3	Sequence 3, Appli
	2	15.4	73.3	12537	3	US-08-611-280-4	Sequence 4, Appli
c	3	15.2	72.4	56516	4	US-08-996-306-1	Sequence 1, Appli
	4	14.8	70.5	1858	4	US-08-909-965C-11	Sequence 11, Appl
c	5	14.8	70.5	1860	2	US-08-624-125-1	Sequence 1, Appli
	6	14.6	69.5	2092	5	US-08-307-896-6	Sequence 6, Appli
	7	14.6	69.5	2092	6	PCT-US95-11808-6	Sequence 6, Appli
	8	14.2	67.6	975	7	5340934-10	Patent No. 5340934
	9	14.2	67.6	977	7	5340934-9	Patent No. 5340934
	10	14.2	67.6	1065	1	US-08-292-549-5	Sequence 5, Appli
	11	14.2	67.6	2924	6	PCT-US93-07923-1	Sequence 1, Appli
	12	14.2	67.6	2974	5	US-08-433-522A-7	Sequence 7, Appli
c	13	13.8	65.7	2368	1	US-07-698-926A-1	Sequence 1, Appli
c	14	13.8	65.7	3114	5	US-08-946-026-12	Sequence 12, Appl
	15	13.8	65.7	4047	4	US-08-612-734B-1	Sequence 1, Appli
	16	13.8	65.7	4800	4	US-08-612-734B-3	Sequence 3, Appli
	17	13.8	65.7	8378	6	PCT-US91-09055-1	Sequence 1, Appli
c	18	13.8	65.7	35100	3	US-08-770-379-19	Sequence 19, Appl
	19	13.6	64.8	1185	1	US-07-876-280-8	Sequence 8, Appli
	20	13.6	64.8	1185	1	US-08-049-783-5	Sequence 5, Appli
	21	13.6	64.8	1185	2	US-08-316-301A-9	Sequence 9, Appli
	22	13.6	64.8	1185	3	US-08-904-278-9	Sequence 9, Appli
	23	13.6	64.8	1185	6	PCT-US92-03624-9	Sequence 9, Appli
c	24	13.6	64.8	1983	1	US-08-073-799C-[illegible]	Sequence 9, Appli
	25	13.6	64.8	2190	5	US-08-508-761B-[illegible]	Sequence 5, Appli
c	26	13.6	64.8	3747	4	US-09-080-897-1	Sequence 1, Appli
c	27	13.6	64.8	3819	1	US-07-686-322A-3	Sequence 3, Appli
c	28	13.6	64.8	3819	1	US-08-002-999-3	Sequence 3, Appli
	29	13.6	64.8	3867	1	US-07-876-280-3	Sequence 3, Appli
	30	13.6	64.8	3867	1	US-07-675-772-3	Sequence 3, Appli
	31	13.6	64.8	3867	1	US-08-063-170-3	Sequence 3, Appli
	32	13.6	64.8	3867	1	US-08-158-232-3	Sequence 3, Appli
	33	13.6	64.8	3867	1	US-08-304-626-3	Sequence 3, Appli
	34	13.6	64.8	3867	2	US-08-316-301A-3	Sequence 3, Appli
	35	13.6	64.8	3867	2	US-08-611-928-3	Sequence 3, Appli
	36	13.6	64.8	3867	6	PCT-US92-03624-3	Sequence 3, Appli
	37	13.6	64.8	3867	7	5281530-4	Patent No. 5281530
	38	13.6	64.8	3867	7	5426049-3	Patent No. 5426049
	39	13.6	64.8	4818	1	US-08-232-463-11	Sequence 11, Appl
c	40	13.6	64.8	5864	5	US-08-894-440-4	Sequence 4, Appli
c	41	13.6	64.8	603[illegible]	2	US-08-441-139-8	Sequence 8, Appli
	42	13.6	64.8	695[illegible]	1	US-07-805-123C-2	Sequence 2, Appli
	43	13.6	64.8	6953	1	US-08-033-081B-2	Sequence 2, Appli
	44	13.4	63.8	51	1	US-08-105-483-209	Sequence 209, App
c	45	13.4	63.8	51	1	US-08-105-483-210	Sequence 210, App

ALIGNMENTS

RESULT 1
US-07-912-015-3/c
; Sequence 3, Application US/07912015
; Patent No. 5283191
; GENERAL INFORMATION:
; APPLICANT: Morgan, Robin Wilson
; APPLICANT: Willemse, Martha Jacoba
; APPLICANT: Claessens, Johannes Antonius Joseph
; APPLICANT: Sondermeijer, Paulus Jacobus Antonius
; TITLE OF INVENTION: Marek's Disease Virus vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/912,015
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/615,211
; FILING DATE: 16-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna Bobrowicz
; REGISTRATION NUMBER: 32,196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301)258-5200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Marek's disease herpesvirus
; STRAIN: GA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 41..3265
; OTHER INFORMATION: /label= pMD20

US-07-912-015-3

Query Match 73.3%; Score 15.4; DB 1; Length 3265;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCG 17
||||||||||||| ||||
Db 1102 AACGAATTCCAAATCCG 1086

RESULT 2
US-08-611-280-4
; Sequence 4, Application US/08611280
; Patent No. 5891666
; GENERAL INFORMATION:
; APPLICANT: Matsuyama, Toshifumi
; APPLICANT: Grossman, Alex
; APPLICANT: Richardson, Christopher D.
; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Canada Inc.
; STREET: 6733 Mississauga Road, Suite 303
; CITY: Mississauga
; STATE: Ontario
; COUNTRY: Canada
; ZIP: L5N 6JB
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,280
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-338A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12537 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-611-280-4

Query Match 73.3%; Score 15.4; DB 3; Length 12537;
Best Local Similarity 94.1%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCG 17
||||||||||||||| |
Db 7719 AACGAATTCCAATTCGG 7735

RESULT 3
US-08-996-306-1/c
; Sequence 1, Application US/08996306
; Patent No. 5945522
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Chumakov, Ilya
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,306
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: GENSET.018A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56516 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: GENOMIC DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Promoter
; LOCATION: 1629..1870
; IDENTIFICATION METHOD: Proscan
; FEATURE:
; NAME/KEY: Potential ATG
; LOCATION: 1998..2000
; FEATURE:
; NAME/KEY: Exon 1
; LOCATION: 2001..2216
; FEATURE:
; NAME/KEY: ATG
; LOCATION: 2031..2033
; FEATURE:
; NAME/KEY: TYR phos
; LOCATION: 11694..14332
; FEATURE:
; NAME/KEY: SEQ ID42
; LOCATION: 11930..11947
; FEATURE:
; NAME/KEY: SEQ ID24
; LOCATION: 12057..12103
; FEATURE:
; NAME/KEY: SEQ ID51
; LOCATION: compl(12339..12358)
; FEATURE:
; NAME/KEY: SEQ ID64
; LOCATION: 13547..13564
; FEATURE:
; NAME/KEY: SEQ ID58
; LOCATION: 13657..13703
; FEATURE:
; NAME/KEY: SEQ ID67
; LOCATION: compl(13962..13981)
; FEATURE:
; NAME/KEY: Exon 2
; LOCATION: 18196..18265
; FEATURE:
; NAME/KEY: Exon 3
; LOCATION: 23717..23832
; FEATURE:
; NAME/KEY: Exon 4

```
;      LOCATION:  25571..25660
;    FEATURE:
;      NAME/KEY:  SEQ ID43
;      LOCATION:  34216..34234
;    FEATURE:
;      NAME/KEY:  SEQ ID25
;      LOCATION:  34469..34515
;    FEATURE:
;      NAME/KEY:  SEQ ID52
;      LOCATION:  compl(34625..34645)
;    FEATURE:
;      NAME/KEY:  Exon 5
;      LOCATION:  34669..34759
;    FEATURE:
;      NAME/KEY:  Exon 6
;      LOCATION:  40688..40846
;    FEATURE:
;      NAME/KEY:  Exon 7
;      LOCATION:  48070..48193
;    FEATURE:
;      NAME/KEY:  Exon 8
;      LOCATION:  50182..54523
;    FEATURE:
;      NAME/KEY:  SEQ ID65
;      LOCATION:  51149..51168
;    FEATURE:
;      NAME/KEY:  SEQ ID59
;      LOCATION:  51448..51494
;    FEATURE:
;      NAME/KEY:  SEQ ID68
;      LOCATION:  compl(51482..51499)
;    FEATURE:
;      NAME/KEY:  SEQ ID44
;      LOCATION:  51596..51613
;    FEATURE:
;      NAME/KEY:  SEQ ID26
;      LOCATION:  51612..51658
;    FEATURE:
;      NAME/KEY:  SEQ ID53
;      LOCATION:  compl(51996..52015)
;    FEATURE:
;      NAME/KEY:  polyAd signal
;      LOCATION:  54445..54450
US-08-996-306-1


  Query Match             72.4%;  Score 15.2;  DB 4;  Length 56516;
  Best Local Similarity   85.0%;  Pred. No. 54;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGTG 21
           | ||||||| |||||||||
Db   12799 AGGAATTCCTGTTCCGGGTG 12780


RESULT   4
US-08-909-965C-11
; Sequence 11, Application US/08909965C
; Patent No. 593607.
;  GENERAL INFORMATION:
;    APPLICANT:  Kuga Tetsro
;    APPLICANT:  Nakagawa Satoshi
;    APPLICANT:  Sakaki yoshiyuki
;    APPLICANT:  Zhao Nanding
;    APPLICANT:  Hashida Hideji
;    TITLE OF INVENTION:  NOVEL DNA, NOVEL POLYPEPTIDE
;    TITLE OF INVENTION:  AND NOVEL ANTIBODY
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITZPATRICK, CELLA, HARPER AND SCINTO
;      STREET:  277 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10172-0194
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/909,965C
;      FILING DATE:  August 12, 1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 322745/95
;      APPLICATION NUMBER:  PCT/JP96/03630
;      FILING DATE:  12-No. 5936078-1995
;      FILING DATE:  12-Dec-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lawrence S. Perry
;      REGISTRATION NUMBER:  31865
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-758-2400
;      TELEFAX:  212-758-2982
;      TELEX:  236262
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;    IMMEDIATE SOURCE:
;      CLONE:  F8232
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  26 to 1255
;      IDENTIFICATION METHOD:  by experiment
US-08-909-965C-11


  Query Match             70.5%;  Score 14.8;  DB 4;  Length 1858;
  Best Local Similarity   88.9%;  Pred. No. 62;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CGAATTCCAATTCCGGGT 20
           ||| |||||| |||||||
Db   1036 CGAGTTCCAAGTCCGGGT 1053


RESULT   5
US-08-624-125-1/c
; Sequence 1, Application US/08624125
; Patent No. 5744341
;  GENERAL INFORMATION:
;    APPLICANT:  CUNNINGHAM JR., FRANCIS X.
;    APPLICANT:  SUN, ZAIREN
;    TITLE OF INVENTION:  GENES OF CAROTENOID BIOSYNTHESIS AND
;    TITLE OF INVENTION:  METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/624,125
;      FILING DATE:  29-MAR-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KELBER, STEVEN B.
;      REGISTRATION NUMBER:  30,073
;      REFERENCE/DOCKET NUMBER:  2747-063-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1860 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  109..1680
;      OTHER INFORMATION:  /product= "E-CYCLASE FROM A.
;      OTHER INFORMATION:  THALIANA"
US-08-624-125-1


  Query Match           70.5%;  Score 14.8;  DB 2;  Length 1860;
  Best Local Similarity  88.9%;  Pred. No. 62;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 GAATTCCAATTCCGGGTG 21
           |||||| ||||||| |||
Db    1159 GAATTCGAATTCCGAGTG 1142


RESULT   6
US-08-307-896-6
; Sequence 6, Application US/08307896C
; Patent No. 6034071
; GENERAL INFORMATION:
;  APPLICANT: Iyengar, Srinivas Ravi
;  TITLE OF INVENTION: MUTANT ACTIVATED GSALPHA AND ADENYLYL
;  TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;  FILE REFERENCE: 29770
;  CURRENT APPLICATION NUMBER: US/08/307,896C
;  CURRENT FILING DATE: 1994-09-16
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2092
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-307-896-6


  Query Match           69.5%;  Score 14.6;  DB 5;  Length 2092;
  Best Local Similarity  81.0%;  Pred. No. 78;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           ||||| ||| |||| || |||
Db    1132 aacgacttcaaattgcgagtg 1152


RESULT   7
PCT-US95-11808-6
; Sequence 6, Application PC/TUS9511808
;  GENERAL INFORMATION:
;    APPLICANT:  Iyengar, Srinivas Ravi V.
;    TITLE OF INVENTION:  MUTANT ACTIVATED GsALPHA AND
;    TITLE OF INVENTION:  ADENYLYL
;    TITLE OF INVENTION:  CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue and
;      ADDRESSEE:  Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11808
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/307,896
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S.
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  29970 165/28755
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 408-2500
;      TELEFAX:  (212) 765-2519
;      TELEX:  650 6111063
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2092 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2092
;      OTHER INFORMATION:  /product= "adenylyl cyclase
;      OTHER INFORMATION:  2
;      OTHER INFORMATION:  (partial)"
PCT-US95-11808-6


  Query Match           69.5%;  Score 14.6;  DB 5;  Length 2092;
  Best Local Similarity  81.0%;  Pred. No. 78;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           ||||| ||| |||| || |||
Db    1132 AACGACTTCAAATTGCGAGTG 1152


RESULT   8
5340934-10
;Patent No. 5340934
;    APPLICANT: TERMINE, JOHN D.;YOUNG, MARIAN F.;FISHER, LARRY W.
; ROBEY, PAMELA G.
;    TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/432,044
;      FILING DATE: 03-NOV-1989
;SEQ ID NO:10:
;      LENGTH: 975
5340934-10


  Query Match           67.6%;  Score 14.2;  DB 7;  Length 975;
  Best Local Similarity  84.2%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGG 19
```

```
          ||||| ||||| ||| |||
Db    176 aacgatttccagttcaggg 194


RESULT   9
5340934-9
;Patent No. 5340934
;    APPLICANT: TERMINE, JOHN D.;YOUNG, MARIAN F.;FISHER, LARRY W.
; ROBEY, PAMELA G.
;    TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/432,044
;      FILING DATE: 03-NOV-1989
;SEQ ID NO:9:
;      LENGTH: 977
5340934-9


  Query Match             67.6%;  Score 14.2;  DB 7;  Length 977;
  Best Local Similarity   84.2%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGG 19
           ||||| ||||| ||| |||
Db    178 aacgatttccagttcaggg 196


RESULT  10
US-08-292-549-5
; Sequence 5, Application US/08292549
; Patent No. 5464938
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Craig A.
;    APPLICANT:  Goodwin, Raymond G.
;    TITLE OF INVENTION:  Isolated Viral Protein TNF Antagonists
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/292,549
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/963,330
;      FILING DATE:  10/19/92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia A.
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  2602-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1065 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Cowpox virus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1065
US-08-292-549-5


  Query Match             67.6%;  Score 14.2;  DB 1;  Length 1065;
  Best Local Similarity   84.2%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGT 20
           || |||||||||||||  ||
Db   1030 ACTAATTCCAATTCCCAGT 1048


RESULT  11
PCT-US93-07923-1
; Sequence 1, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfec  (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2924
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US93-07923-1


  Query Match             67.6%;  Score 14.2;  DB 6;  Length 2924;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGG 19
           |||||||| |||||  |||
Db   1875 AACGAATTGCAATTTGGGG 1893


RESULT  12
```

```
US-08-433-522A-7
; Sequence 7, Application US/08433522A
; Patent No. 6013514
;  GENERAL INFORMATION:
;    APPLICANT:  CHONG, Pele
;    APPLICANT:  THOMAS, Wayne
;    APPLICANT:  YANG, Yan Ping
;    APPLICANT:  LOOSMORE, Sheena
;    APPLICANT:  SIA, Dwo Yuan Charles
;    APPLICANT:  KLEIN, Michel
;    TITLE OF INVENTION:  HAEMOPHILUS OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6TH Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,522A
;      FILING DATE:  12-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-434 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2974 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  386..2761
US-08-433-522A-7


  Query Match             67.6%;  Score 14.2;  DB 5;  Length 2974;
  Best Local Similarity   84.2%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCAATTCCGGGT 20
           |||| || ||||||| |||
Db    2152 ACGAGTTACAATTCCAGGT 2170


RESULT  13
US-07-698-926A-1/c
; Sequence 1, Application US/07698926A
; Patent No. 5426052
;  GENERAL INFORMATION:
;    APPLICANT:  Flickinger, Michael C.
;    APPLICANT:  Mills, David A.
;    TITLE OF INVENTION:  Bacillus MGA3 Diaminopimelate
;    TITLE OF INVENTION:  Decarboxylase Gene
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  3100 No. 5426052west Center
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/698,926A
;      FILING DATE:  19910510
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Woessner, Warren D
;      REGISTRATION NUMBER:  30,440
;      REFERENCE/DOCKET NUMBER:  600.217-US-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-332-5300
;      TELEFAX:  612-332-9081
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2368 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  DAP Decarboxylase Gene
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  265..1560
US-07-698-926A-1


  Query Match             65.7%;  Score 13.8;  DB 1;  Length 2368;
  Best Local Similarity   88.2%;  Pred. No. 2e+02;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCG 17
           |||||||||| | ||||
Db    1021 AACGAATTCCGAATCCG 1 )5


RESULT  14
US-08-946-026-12/c
; Sequence 12, Application US/08946026
; Patent No. 6034218
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Mitcham, Jennifer L.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
;    TITLE OF INVENTION:  AND IMMUNODIAGNOSIS OF PROSTATE CANCER
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,026
;      FILING DATE:  07-OCT-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.424C1
```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3114 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-946-026-12

Query Match 65.7%; Score 13.8; DB 5; Length 3114;
Best Local Similarity 88.2%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GAATTCCAATTCCGGGT 20
|||||| |||||| |||
Db 2638 GAATTCGAATTCCAGGT 2622

RESULT 15
US-08-612-734B-1
; Sequence 1, Application US/08612734B
; Patent No. 5914246
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Tobin, Matthew B.
; TITLE OF INVENTION: Multiple Drug Resistance Gene of
; TITLE OF INVENTION: Aspergillus Fumigatus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center, DC1501
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,734B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Craig, Anne I.
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: X-9681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-354-9570
; TELEFAX: 617-354-4043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4047
US-08-612-734B-1

Query Match 65.7%; Score 13.8; DB 4; Length 4047;
Best Local Similarity 88.2%; Pred. No. 2.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CGAATTCCAATTCCGGG 19
|||| || |||||||||
Db 3306 CGAAATCGAATTCCGGG 3322

Search completed: June 22, 2000, 18:04:14
Job time: 9721 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:45 ; Search time 2113.77 Seconds
(without alignments)
40.268 Million cell updates/sec

Title: US-09-362-485-26
Perfect score: 21
Sequence: 1 AACGAATTCCAATTCCGGGTG 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
c	1	17.6	83.8	401	107	AQ444869	AQ444869 GSSTc0453
c	2	16.8	80.0	168	22	R57765	R57765 F5460 Fetal
	3	16.4	78.1	557	109	AQ609640	AQ609640 HS_5079_A
c	4	16.2	77.1	219	47	AI496177	AI496177 sa96c09.y
c	5	16.2	77.1	258	64	AW055881	AW055881 SWYACAL01
c	6	16.2	77.1	283	73	AV343627	AV343627 AV343627
c	7	16.2	77.1	311	45	AI350859	AI350859 qt11f12.x
	8	16.2	77.1	360	36	C62596	C62596 C62596 Yuji
	9	16.2	77.1	373	36	C65178	C65178 C65178 Yuji
	10	16.2	77.1	374	38	AA758291	AA758291 ah65b06.s
c	11	16.2	77.1	455	33	AA406933	AA406933 MBAFCZ7A0
	12	16.2	77.1	458	38	AA770303	AA770303 ah79d10.s
	13	16.2	77.1	476	28	AA110821	AA110821 mm89e02.r
c	14	16.2	77.1	537	105	AQ270237	AQ270237 HS_2045_A
c	15	16.2	77.1	1101	83	CNS014QL	AL104487 Drosophil
	16	16	76.2	395	40	AA961280	AA961280 on96c02.s
	17	16	76.2	405	37	AA723779	AA723779 ah87a06.s
	18	16	76.2	435	45	AI393115	AI393115 tg41b01.x
	19	16	76.2	898	39	AA890720	AA890720 ak13d01.s
	20	16	76.2	1031	82	CNS006Y0	AL066067 Drosophil
c	21	15.8	75.2	262	70	AV238441	AV238441 AV238441
	22	15.8	75.2	368	62	AU076018	AU076018 AU076018
c	23	15.8	75.2	399	33	AA406659	AA406659 MBAFCZ6B1
c	24	15.8	75.2	481	34	AA463551	AA463551 zx72c10.r
	25	15.8	75.2	484	91	AQ857146	AQ857146 nbeb0005E
	26	15.8	75.2	491	39	AA865695	AA865695 og95g05.s
	27	15.8	75.2	542	106	AQ394543	AQ394543 CITBI-E1-
c	28	15.8	75.2	655	92	AQ945264	AQ945264 Sheared D
	29	15.8	75.2	702	108	AQ574869	AQ574869 nbxb0086A
c	30	15.8	75.2	773	94	AQ050478	AQ050478 nbxb0004a
c	31	15.8	75.2	1101	82	CNS00E7O	AL069047 Drosophil
c	32	15.4	73.3	199	71	AV316899	AV316899 AV316899
	33	15.4	73.3	247	37	AA688418	AA688418 nv59b10.s
c	34	15.4	73.3	300	35	C31642	C31642 C31642 Yuji
c	35	15.4	73.3	300	35	C53491	C53491 C53491 Yuji
c	36	15.4	73.3	300	35	C56806	C56806 C56806 Yuji
c	37	15.4	73.3	300	62	AV182768	AV182768 AV182768
c	38	15.4	73.3	325	24	D73316	D73316 CELK117CYR
c	39	15.4	73.3	326	24	D64792	D64792 CELK088G2R
	40	15.4	73.3	338	36	AA644089	AA644089 ab62b03.s
c	41	15.4	73.3	395	46	AI442551	AI442551 sa32h10.y
c	42	15.4	73.3	407	74	AW206302	AW206302 UI-H-BI1-
	43	15.4	73.3	438	105	AQ304222	AQ304222 HS_3213_A
c	44	15.4	73.3	510	102	AQ157706	AQ157706 nbxb0009P
	45	15.4	73.3	526	85	AQ718130	AQ718130 HS_5511_B

ALIGNMENTS

RESULT 1
AQ444869/c
LOCUS AQ444869 401 bp DNA GSS 31-MAR-1999
DEFINITION GSSTc0453 Trypanosoma cruzi random genomic library Trypanosoma cruzi genomic clone G11020 5', genomic survey sequence.
ACCESSION AQ444869
VERSION AQ444869.1 GI:4556333
KEYWORDS GSS.
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 401)
AUTHORS Sanchez,D.O.
TITLE Trypanosoma cruzi random genomic sequences
JOURNAL Unpublished (1999)
COMMENT Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral San Martin)
Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
CP(1650) San Martin, Prov. de BS AS. Argentina
Tel: (54-11)4752-0021
Fax: (54-11)4752-9639
Email: dsanchez@inti.gov.ar
Seq primer: T7
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .401
/organism="Trypanosoma cruzi"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="G11020"
/clone_lib="Trypanosoma cruzi random genomic library"
/cell_type="epimastigote"
/note="Vector: pBS(-) (PHARMACIA)"
BASE COUNT 95 a 88 c 131 g 76 t 11 others
ORIGIN

Query Match 83.8%; Score 17.6; DB 107; Length 401;
Best Local Similarity 81.0%; Pred. No. 35;
Matches 17; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 AACGAATTCCAATTCCGGGTG 21
          |: |||||||||:|||||| |
Db    398 AWTGAATTCCAAWTCCGGGNG 378
```

RESULT 2
R57765/c
LOCUS R57765 168 bp mRNA EST 02-MAY-1996
DEFINITION F5460 Fetal heart Homo sapiens cDNA clone F5460 5' end, mRNA sequence.
ACCESSION R57765
VERSION R57765.1 GI:827823
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168)
AUTHORS Hwang,D.M., Fung,Y.W., Wang,R.X., Laurenssen,C.M., Ng,S.H., Lam,W.Y., Tsui,K.W., Fung,K.P., Waye,M., Lee,C.Y. and Liew,C.C.
TITLE Analysis of expressed sequence tags from a fetal human heart cDNA library
JOURNAL Genomics 30 (2), 293-298 (1995)
MEDLINE 96163883
COMMENT Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Insert Length: 1535 Std Error: 0.00
Seq primer: GGTGGCGACGACTCCTGGAGCC
High quality sequence stop: 298.
FEATURES Location/Qualifiers
source 1. .168
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="F5460"
/clone_lib="Fetal heart"
/lab_host="E. coli Y1090"
/note="Vector: Lambda gt22; Site_1: NotI; Site_2: SalI;
mRNA was purified from human fetal hearts (10-12 weeks).
cDNA was constructed using a NotI-Oligo dT adaptor-primer.
SalI adaptors were ligated, followed by digestion with
NotI, for direction cloning into predigested lambda gt22.
Method is described in J. Mol. Cell. Cardiol. (1994) 26,
1329-1333)"
BASE COUNT 52 a 34 c 37 g 45 t
ORIGIN

```
  Query Match             80.0%;  Score 16.8;  DB 22;  Length 168;
  Best Local Similarity   90.0%;  Pred. No. 83;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGT 20
           |||||||||  ||||  |||||
Db     164 AACGAATTCTAATTGCGGGT 145


RESULT   3
AQ609640
LOCUS       AQ609640      557 bp    DNA             GSS       15-JUN-1999
DEFINITION  HS_5079_A2_F08_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=655 Col=16 Row=K, genomic survey sequence.
ACCESSION   AQ609640
VERSION     AQ609640.1  GI:5070916
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 655  row: K  column: 16
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 557.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=655 Col=16 Row=K"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT      170 a    104 c    131 g    118 t     34 others
ORIGIN


  Query Match             78.1%;  Score 16.4;  DB 109;  Length 557;
  Best Local Similarity   94.4%;  Pred. No. 1.5e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAATTCCAATTCCGGGTG 21
           |||||||||||||  ||||
Db     187 GAATTCCAATTCCTGGTG 204


RESULT   4
AI496177/c
LOCUS       AI496177      219 bp    mRNA            EST       01-DEC-1999
DEFINITION  sa96c09.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1004-7145 5', mRNA sequence.
ACCESSION   AI496177
VERSION     AI496177.1  GI:4397180
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 219)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188340.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 212.
FEATURES             Location/Qualifiers
     source          1. .219
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-7145"
                     /clone_lib="Gm-c1004"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Root cDNA. The mRNA was isolated from entire roots
                     of 8 day old 'Williams' seedlings which were propagated on
                     paper towels with distilled water. Stratagene's cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize the
                     cDNA. First- strand synthesis was performed with 5-methyl
                     dCTP, hence the ligated cDNA is hemimethylated.
                     Stratagene's first-strand synthesis primer was used
                     [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                     second-strand synthesis, the cDNA ends were 'polished'
                     with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                     and phosphorylated. The XhoI site within the first-strand
                     synthesis primer was restricted by digestion with XhoI;
                     all XhoI sites in the  cDNA would be protected by their
                     hemimethylated status. The cDNA constructs were
                     size-fractionated with a 500bp cutoff, using GibcoBRL Life
                     Technologies' cDNA Size Fractionation column. The column
                     eluent was then ligated into  Stratagene's pBluescript II
                     XR Predigested vector (pBluescript II SK(+) that had been
                     digested with EcoRI and XhoI, and phosphorylated). Both
                     the white  and blue colonies appear to contain recombinant
                     plasmids with cDNA inserts. Blue colonies 9n=15) have been
                     sequenced, and possess putative cDNA inserts. This library
                     was constructed by Dr. Paul Keim & Virginia H. Coryell,
                     Department of Biology, Box5640, Northern Arizona
                     University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
                     Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
                     520-523-7500, email: paul.keim@nau.edu,
                     virginia.coryell@nau.edu"
BASE COUNT       76 a     46 c     52 g     45 t
ORIGIN
```

```
  Query Match             77.1%;  Score 16.2;  DB 47;  Length 219;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           ||| ||||||| ||| |||||
Db      41 AACCAATTCCATTTCTGGGTG 21


RESULT   5
AW055881/c
LOCUS       AW055881      258 bp    mRNA            EST       24-SEP-1999
DEFINITION  SWYACAL01H12SK Brugia malayi young adult cDNA (SAW99MLW-BmYA)
            Brugia malayi cDNA clone SWYACAL01H12 5', mRNA sequence.
ACCESSION   AW055881
VERSION     AW055881.1  GI:5922592
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   Williams,S.A.
  TITLE     Genes expressed in young adult of Brugia malayi
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5407310.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .258
                     /organism="Brugia malayi"
                     /db_xref="taxon:6279"
                     /clone="SWYACAL01H12"
                     /clone_lib="Brugia malayi young adult cDNA
                     (SAW99MLW-BmYA)"
                     /dev_stage="young adult"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from young adult worms isolated from the
                     peritoneal cavity of jirds and converted to
                     double-stranded cDNA using reverse transcriptase and
                     oligo(dT) followed by RNase H and DNA pol I. The library
                     has 6.5 x 10^4 independent recombinants and the average
                     insert size is approx. 800bp.  The library was constructed
                     by Michelle Lizotte-Waniewski.  The library is available
                     from Dr. S.A. Williams,  email: genome@neal.smith.edu."
BASE COUNT       78 a     62 c     51 g     65 t      2 others
ORIGIN


  Query Match             77.1%;  Score 16.2;  DB 64;  Length 258;
  Best Local Similarity   85.7%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           |||||||||||  || ||||
Db     174 AACGAATTCCAAAACCTGGTG 154


RESULT   6
AV343627/c
LOCUS       AV343627      283 bp    mRNA            EST       12-NOV-1999
DEFINITION  AV343627 RIKEN full-length enriched, adult male olfactory brain Mus
            musculus cDNA clone 6430538M11 3' similar to AF030358 Rattus
            norvegicus chemokine CX3C mRNA, mRNA sequence.
ACCESSION   AV343627
VERSION     AV343627.1  GI:6384686
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J.  Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Mii ,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 9, 1999 this sequence version replaced gi:5435482.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6430538M11"
                     /clone_lib="RIKEN full-length enriched, adult male
                     olfactory brain"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
```

3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 62 a 71 c 43 g 107 t
ORIGIN

Query Match 77.1%; Score 16.2; DB 73; Length 283;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||| |||||||||| |||||
Db 131 AACAAATTCCAATTATGGGTG 111

RESULT 7
AI350859/c
LOCUS AI350859 311 bp mRNA EST 13-FEB-1999
DEFINITION qt11f12.x1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1947311 3' similar to SW:COX2_HUMAN P00403 CYTOCHROME C OXIDASE POLYPEPTIDE II ;, mRNA sequence.
ACCESSION AI350859
VERSION AI350859.1 GI:4088065
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:803093.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 354 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 108.
FEATURES Location/Qualifiers
source 1. .311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1947311"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 80 a 47 c 79 g 105 t
ORIGIN

Query Match 77.1%; Score 16.2; DB 45; Length 311;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||| || |||||||||||| |
Db 232 AACCAAATCCAATTCCGGGAG 212

RESULT 8
C62596
LOCUS C62596 360 bp mRNA EST 22-SEP-1997
DEFINITION C62596 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk288c2 5', mRNA sequence.
ACCESSION C62596
VERSION C62596.1 GI:2421301
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Jan 24, 1995 this sequence version replaced gi:634314.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk288c2"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 102 a 89 c 71 g 96 t 2 others
ORIGIN

Query Match 77.1%; Score 16.2; DB 35; Length 360;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||| |||||| |||||| |||
Db 128 AACAAATTCCCATTCCGTGTG 148

RESULT 9
C65178
LOCUS C65178 373 bp mRNA EST 22-SEP-1997
DEFINITION C65178 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk406d11 5', mRNA sequence.
ACCESSION C65178
VERSION C65178.1 GI:2423883
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 373)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397430.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES Location/Qualifiers
source 1. .373
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk406d11"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 107 a 92 c 73 g 100 t 1 others
ORIGIN

Query Match 77.1%; Score 16.2; DB 36; Length 373;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||| |||||| |||||| |||
Db 137 AACAAATTCCCATTCCGTGTG 157

RESULT 10
AA758291
LOCUS AA758291 374 bp mRNA EST 29-DEC-1998
DEFINITION ah65b06.s1 Soares_testis_NHT Homo sapiens cDNA clone 1320467 3', mRNA sequence.
ACCESSION AA758291
VERSION AA758291.1 GI:2806154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 374)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2284745.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 587 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 362.

FEATURES Location/Qualifiers
source 1. .374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1320467"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 134 a 51 c 71 g 118 t
ORIGIN

Query Match 77.1%; Score 16.2; DB 38; Length 374;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGAATTCCAATTCCGGGTG 21
||||| | ||||||| |||||
Db 262 AACGACTGCCAATTCAGGGTG 282

RESULT 11
AA406933/c
LOCUS AA406933 455 bp mRNA EST 01-MAY-1997
DEFINITION MBAFCZ7A04T3 Brugia malayi adult female cDNA (SAW96MLW-BmAF) Brugia malayi cDNA clone AFCZ7A04 5', mRNA sequence.
ACCESSION AA406933
VERSION AA406933.1 GI:2064925
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 455)
AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones,S.J.
TITLE Genes expressed in adult female Brugia malayi
JOURNAL Unpublished (1996)
COMMENT On Jan 24, 1995 this sequence version replaced gi:634308.
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
http://www.sanger.ac.uk/brugia/AFC/MBAFCZ7A04T3.html
Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .455
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="AFCZ7A04"
/clone_lib="Brugia malayi adult female cDNA (SAW96MLW-BmAF)"
/sex="female"
/dev_stage="adult"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from approximately 50 adult females isolated from the peritoneal cavity of jirds and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 5 x 10E6 independent recombinants and the average insert size is ~900bp.The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email: genome@smith.edu."

BASE COUNT 135 a 98 c 95 g 126 t 1 others
ORIGIN

```
  Query Match            77.1%;  Score 16.2;  DB 33;  Length 455;
  Best Local Similarity  85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          ||||||||||||  || ||||
Db    172 AACGAATTCCAAAACCTGGTG 152


RESULT  12
AA770303
LOCUS       AA770303      458 bp    mRNA            EST       29-DEC-1998
DEFINITION  ah79d10.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321843 3',
            mRNA sequence.
ACCESSION   AA770303
VERSION     AA770303.1  GI:2821541
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:934360.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 579   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="1321843"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      163 a     68 c     87 g    140 t
ORIGIN


  Query Match            77.1%;  Score 16.2;  DB 38;  Length 458;
  Best Local Similarity  85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          ||||| | |||||||  |||||
Db    263 AACGACTGCCAATTCAGGGTG 283


RESULT  13
AA110821
LOCUS       AA110821      476 bp    mRNA            EST       07-FEB-1997
DEFINITION  mm89e02.r1 Stratagene mouse embryonic carcinomaRA (#937318) Mus
            musculus cDNA clone IMAGE:535610 5' similar to gb:X54802 CYTOCHROME
            C OXIDASE POLYPEPTIDE IV PRECURSOR (HUMAN); gb:M37831 Mouse
            mitochondrial cytochrome oxidase subunit IV (MOUSE);, mRNA
            sequence.
ACCESSION   AA110821
VERSION     AA110821.1  GI:1662716
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:322546
            Trace considered overall poor quality
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:535610"
                     /clone_lib="Stratagene mouse embryonic carcinomaRA
                     (#937318)"
                     /tissue_type="carcinoma"
                     /dev_stage="embryonic"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. P19 cell
                     line treated with retinoic acid. Average insert size: 1.0
                     kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      129 a     110 c     105 g    132 t
ORIGIN


  Query Match            77.1%;  Score 16.2;  DB 28;  Length 476;
  Best Local Similarity  85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCAATTCCGGGTG 21
          | | ||||||||||||| |||
Db     38 AGCCAATTCCAATTCCGTGTG 58


RESULT  14
AQ270237/c
LOCUS       AQ270237      537 bp    DNA             GSS       03-NOV-1998
DEFINITION  HS_2045_A1_F01_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2045 Col=1 Row=K, genomic survey
            sequence.
```

```
ACCESSION   AQ270237
VERSION     AQ270237.1  GI:3822832
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2045  row: K  column: 1
            Class: BAC ends
            High quality sequence stop: 537.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2045 Col=1 Row=K"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      140 a    128 c    108 g    145 t     16 others
ORIGIN
```

```
  Query Match             77.1%;  Score 16.2;  DB 105;  Length 537;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           | ||||| |||||||| ||||
Db     452 ATCGAATACCAATTCCAGGTG 432
```

```
RESULT  15
CNS014QL/c
LOCUS       CNS014QL     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12L15 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104487
VERSION     AL104487.1  GI:5616258
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN12L15"
                     /note="end : SP6"
BASE COUNT      336 a    213 c    165 g    308 t     79 others
ORIGIN
```

```
  Query Match             77.1%;  Score 16.2;  DB 83;  Length 1101;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCAATTCCGGGTG 21
           || ||| || ||||||||||||
Db     567 AAAGAACTCAAATTCCGGGTG 547
```

```
Search completed: June 22, 2000, 17:43:49
Job time: 20099 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 17:43:56 ; Search time 40.17 Seconds
(without alignments)
4.128 Million cell updates/sec

Title: US-09-362-485-27
Perfect score: 37
Sequence: 1 NEFQFRV 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	28	75.7	10	1	R96041	Telomerase peptide
2	28	75.7	22	1	W97830	Human titin fibron
3	28	75.7	54	1	W79439	Staphylococcus aur
4	28	75.7	146	1	W79438	Staphylococcus aur
5	28	75.7	157	1	W20538	H. pylori outer me
6	28	75.7	253	1	W20760	H. pylori outer me
7	28	75.7	355	1	W98871	H. pylori GHPO 172
8	28	75.7	356	1	Y10960	H. pylori ORF 07ap
9	28	75.7	409	1	R90577	Mouse TRAF1. Tumou
10	28	75.7	443	1	Y05193	S. aureus priA pro
11	28	75.7	624	1	W14849	Enzyme which decom
12	28	75.7	780	1	Y05192	S. aureus priA pro
13	28	75.7	872	1	R96039	95 kD protein. Tel
14	27	73.0	76	1	Y12640	Human 5' EST secre
15	27	73.0	539	1	W62301	Spinach D1 proteas
16	26	70.3	184	1	W90344	S. mutans lanB pro
17	26	70.3	403	1	W68561	S. pneumoniae dnrX
18	26	70.3	802	1	R44242	F-spondin of chick
19	25	67.6	31	1	Y11250	Streptococcus pneu
20	25	67.6	74	1	Y12623	Human 5' EST secre
21	25	67.6	96	1	W41413	PreproHCPB fragmen
22	25	67.6	96	1	W74477	Pro-L amino acid s
23	25	67.6	99	1	W74478	pro-KDEL amino aci
24	25	67.6	99	1	W74479	Pro-KKAA amino aci
25	25	67.6	101	1	W74480	Pro-SDYQRL amino a
26	25	67.6	118	1	R07611	Murine granulocyte
27	25	67.6	118	1	P50294	Sequence of mammal
28	25	67.6	141	1	R92799	Mouse GM-CSF. Mute
29	25	67.6	141	1	W56712	Murine granulocyte
30	25	67.6	232	1	Y00874	pHOOK-1.GM-CSF fus
31	25	67.6	255	1	W12778	Japanese horseshoe
32	25	67.6	400	1	R56521	Protein kirase (Hh
33	25	67.6	400	1	R76617	Schizosaccharomyce
34	25	67.6	415	1	W06172	Human pancreatic c
35	25	67.6	415	1	W74476	Human pancreatic c
36	25	67.6	424	1	W06175	ProHCPB with PelB
37	25	67.6	438	1	W80756	Mouse cytokine rec
38	25	67.6	442	1	R77259	Exo-cellobiohydrol
39	25	67.6	442	1	W80754	Human cytokine rec
40	25	67.6	458	1	W80755	Mouse cytokine rec
41	25	67.6	510	1	W38289	Murine N-acetylgal
42	25	67.6	513	1	P50133	Sequence encoded b
43	25	67.6	513	1	R77261	T. longibrachiatum
44	25	67.6	513	1	W02022	Trichoderma cellob
45	25	67.6	514	1	W57419	Regulatory sequenc

ALIGNMENTS

RESULT 1
R96041
ID R96041 standard; peptide; 10 AA.
AC R96041;
DT 24-FEB-1997 (first entry)
DE Telomerase peptide fragment #2.
KW Tetrahymena; 80 kD protein; 95 kD protein; telomerase; telomeric repeat;
KW RNA coupled protein; chromosome; telomere repeat synthesis; antibody;
KW immortalised cell; cancer; eukaryotic microbe; inhibitor; gene therapy;
KW fungal infection; therapy; diagnosis; protozoan infection; somatic cell;
KW p80; p95.
OS Synthetic.
PN WO9619580-A2.
PD 27-JUN-1996.
PF 18-DEC-1995; U16531.
PR 19-DEC-1994; US-359125.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Autexier C, Collins K, Greider C, Hemish JM, Kobayashi R;
PI Yang XH;
DR WPI; 96-309594/31.
PT Telomerase protein and related DNA, antibodies, transgenic cells,
PT etc. - for diagnosis and treatment of cancer and infection by
PT eukaryotic microbes, also new telomerase inhibitors
PS Claim 11; Page 39; 56pp; English.
CC R96040-R96042 represent peptide fragments of telomerase. These sequences
CC are recognised by antibodies which bind to the 80 kD and 95 kD proteins
CC of Tetrahymena (see R96038 and R96039 respectively). The 80 kD and 95 kD
CC proteins are components of telomerase. The proteins act, when coupled to
CC RNA, to add telomeric repeats (of the sequence TTGGGG) to stabilise
CC chromosomes. The RNA component provides the template for the telomere
CC repeat synthesis. Antibodies against these sequences, can be used as
CC immunoassay reagents for the detection of the proteins. The antibodies
CC can also be used to identify immortalised cells, or predisposition to
CC immortalisation, particularly cancer, or to diagnose a disease caused by
CC a eukaryotic microbe. Inhibitors of the 80 kD and 95 kD proteins (and
CC the proteins themselves) can be used for therapy or diagnosis. The
CC inhibitors can be used to treat infection by fungi and protozoa. As
CC somatic cells do not generally require telomerase, these inhibitors
CC should have little or no toxicity to the host. The DNA encoding the 80
CC kD and 95 kD sequences are used to produce the recombinant protein, or to
CC isolate similar genes from other organisms, while transformed cells can
CC be used in gene therapy.
SQ Sequence 10 AA;

Query Match 75.7%; Score 28; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQF 5
|||||
Db 2 NEFQF 6

RESULT 2
W97830
ID W97830 standard; Peptide; 22 AA.

```
AC   W97830;
DT   21-MAY-1999  (first entry)
DE   Human titin fibronectin repeat peptide.
KW   Swollenin; cellulose; textile; feedstuff; paper; titin; human.
OS   Homo sapiens.
PN   WO9902693-A2.
PD   21-JAN-1999.
PF   09-JUL-1998; U14226.
PR   11-JUL-1997; US-893766.
PA   (GEMV ) GENENCOR INT INC.
PI   Penttila M, Pere J, Saloheimo M, Swanson BA, Ward M;
DR   WPI; 99-120894/10.
PT   Novel microbial swollenin protein - useful to alter the cellulosic
PT   properties of a substrate e.g. in animal nutrition, paper making,
PT   laundry and textile detergents, etc
PS   Example 1; Page 17; 39pp; English.
CC   This peptide comprises amino acids 12268-89 of human titin, in
CC   a fibronectin repeat region.  It shows homology to a region (see
CC   W97829) of the novel swollenin protein (see W97823) of Trichoderma
CC   reesei.  Swollenin has the ability to weaken filter paper and
CC   swell cellulose without having cellulolytic activity.  It can be
CC   used for altering the cellulosic properties of a substrate, e.g.
CC   in animal feeds, paper making, laundry and textile detergents, etc.
SQ   Sequence   22 AA;

  Query Match             75.7%;  Score 28;  DB 1;  Length 22;
  Best Local Similarity   71.4%;  Pred. No. 4.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||: |||
Db       2 NEYYFRV 8


RESULT   3
W79439
ID   W79439 standard; Protein; 54 AA.
AC   W79439;
DT   27-NOV-1998  (first entry)
DE   Staphylococcus aureus YqkL protein.
KW   Bacillus subtilis; YqkL; Neisseria meningitidis; Fur protein;
KW   ferric uptake regulator; treatment; prevention; bacterial infection;
KW   Helicobacter pylori; vaccine.
OS   Staphylococcus aureus.
PN   WO9823738-A2.
PD   04-JUN-1998.
PF   24-NOV-1997; U22092.
PR   25-NOV-1996; US-031469.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Warren RL;
DR   WPI; 98-322718/28.
DR   N-PSDB; V59913.
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
PS   Claim 5; Page 93; 114pp; English.
CC   W79438-40 represent Staphylococcus aureus WCHU (NCIMB
CC   40771) proteins that have homology to a Bacillus subtilis YqkL
CC   protein (W79438) and a Neisseria meningitidis Fur protein and ferric
CC   uptake regulator (W79439-40). The S. aureus proteins are used to
CC   generate antibodies and to screen for antimicrobials. The products
CC   are used to treat or prevent bacterial infections, particularly
CC   where caused by S. aureus but also against Helicobacter pylori.
CC   Particular applications are to treat subjects before surgery or
CC   insertion of an in-dwelling device (alternatively the device itself
CC   is impregnated before placement). The nucleic acid sequence is used
CC   as sources of antisense sequences (for therapeutic use) or
CC   regulatory elements for controlling expression of bacterial genes,
CC   and for antibacterial screening. The protein can be also used as a
CC   vaccine.
SQ   Sequence   54 AA;

  Query Match             75.7%;  Score 28;  DB 1;  Length 54;
  Best Local Similarity   57.1%;  Pred. No. 13;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||| |::
Db      29 NEFNFKI 35


RESULT   4
W79438
ID   W79438 standard; Protein; 146 AA.
AC   W79438;
DT   27-NOV-1998  (first entry)
DE   Staphylococcus aureus YqkL protein.
KW   Bacillus subtilis; YqkL; Neisseria meningitidis; Fur protein;
KW   ferric uptake regulator; treatment; prevention; bacterial infection;
KW   Helicobacter pylori; vaccine.
OS   Staphylococcus aureus.
PN   WO9823738-A2.
PD   04-JUN-1998.
PF   24-NOV-1997; U22092.
PR   25-NOV-1996; US-031469.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Warren RL;
DR   WPI; 98-322718/28.
DR   N-PSDB; V59913.
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
PS   Claim 5; Page 92; 114pp; English.
CC   W79438-40 represent Staphylococcus aureus WCHU (NCIMB
CC   40771) proteins that have homology to a Bacillus subtilis YqkL
CC   protein (W79438) and a Neisseria meningitidis Fur protein and ferric
CC   uptake regulator (W79439-40). The S. aureus proteins are used to
CC   generate antibodies and to screen for antimicrobials. The products
CC   are used to treat or prevent bacterial infections, particularly
CC   where caused by S. aureus but also against Helicobacter pylori.
CC   Particular applications are to treat subjects before surgery or
CC   insertion of an in-dwelling device (alternatively the device itself
CC   is impregnated before placement). The nucleic acid sequence is used
CC   as sources of antisense sequences (for therapeutic use) or
CC   regulatory elements for controlling expression of bacterial genes,
CC   and for antibacterial screening. The protein can be also used as a
CC   vaccine.
SQ   Sequence   146 AA;

  Query Match             75.7%;  Score 28;  DB 1;  Length 146;
  Best Local Similarity   57.1%;  Pred. No. 36;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||| |::
Db     123 NEFNFKI 129


RESULT   5
W20538
ID   W20538 standard; Protein; 157 AA.
AC   W20538;
DT   15-JUL-1997  (first entry)
DE   H. pylori outer membrane protein 5083193.aa.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   misc_difference 104
FT                   /note= "encoded by YAA"
FT   misc_difference 132
FT                   /note= "encoded by GYT"
```

PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
DR WPI; 97-052306/05.
DR N-PSDB; T67689.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Page 697; 1481pp; English.
CC This sequence is a H. pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 157 AA;

Query Match 75.7%; Score 28; DB 1; Length 157;
Best Local Similarity 57.1%; Pred. No. 39;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||||:|::
Db 106 NEFKFQI 112

RESULT 6
W20760
ID W20760 standard; protein; 253 AA.
AC W20760;
DT 15-JUL-1997 (first entry)
DE H. pylori outer membrane protein, 07ap80601orf8.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
KW outer membrane.
OS Helicobacter pylori.
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
DR WPI; 97-052306/05.
DR N-PSDB; T68013.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Page 1173-1173; 1481pp; English.
CC The present sequence is a Helicobacter pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 253 AA;

Query Match 75.7%; Score 28; DB 1; Length 253;
Best Local Similarity 57.1%; Pred. No. 64;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||||:|::
Db 109 NEFKFQI 115

RESULT 7
W98871
ID W98871 standard; Protein; 355 AA.
AC W98871;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 1723 protein.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
OS Helicobacter pylori.
PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR N-PSDB; X14590.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 1976-1977; 2054pp; English.
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 355 AA;

Query Match 75.7%; Score 28; DB 1; Length 355;
Best Local Similarity 57.1%; Pred. No. 91;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||||:|::
Db 106 NEFKFQI 112

RESULT 8
Y10960
ID Y10960 standard; Protein; 356 AA.
AC Y10960;
DT 08-JUN-1999 (first entry)
DE H. pylori ORF 07ap80601_5083193_f3_8 cell envelope protein.
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW secreted protein; cellular protein.
OS Helicobacter pylori.
PN WO9818323-A1.
PD 07-MAY-1998.
PF 28-OCT-1997; U19575.
PR 14-JUL-1997; US-891928.
PR 28-OCT-1996; US-739150.

```
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30427.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 160-161; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   356 AA;

  Query Match             75.7%;  Score 28;  DB 1   Length 356;
  Best Local Similarity   57.1%;  Pred. No. 92;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||:|::
Db     107 NEFKFQI 113


RESULT   9
R90577
ID   R90577 standard; Protein; 409 AA.
AC   R90577;
DT   09-APR-1996  (first entry)
DE   Mouse TRAF1.
KW   TRAF1; tumour necrosis factor receptor associated factor 1;
KW   TNF; CD40.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   domain          180. .409
FT                   /label= TRAF_domain
FT   region          183. .259
FT                   /label= Leucine_zipper_region
PN   WO9533051-A1.
PD   07-DEC-1995.
PF   25-MAY-1995; U06639.
PR   27-MAY-1994; US-250858.
PR   28-OCT-1994; US-331394.
PR   22-MAY-1995; US-446915.
PA   (GETH ) GENENTECH INC.
PI   Goeddel DV,  Rothe M;
DR   WPI; 96-049310/05.
DR   N-PSDB; T12261.
PT   Tumour necrosis factor (TNF) receptor-associated factors - involved
PT   in mediation of biological activities of TNF and CD40 ligands
PS   Claim 8; Page 71-72; 116pp; English.
CC   Mouse tumour necrosis factor receptor associated factor 1 (TRAF1)
CC   (R90577) is a new factor capable of specific association with the
CC   intracellular domain of the type 2 TNF receptor (TNF-R2) and CD40,
CC   and is involved in the mediation of TNF and CD40 ligand biological
CC   activities.  Recombinant TRAF1 is obtd. by expression in host cells
CC   of a cDNA clone (T12261) isolated from murine interleukin-2-dependent
CC   cytotoxic T-cell line CT6.  It is used to identify inhibitors of
CC   activities of TNF-R2, CD40 and/or LMP1 oncogene, e.g. for treratment
CC   of endotoxic (septic) shock and rheumatoid arthritis.
SQ   Sequence   409 AA;

  Query Match             75.7%;  Score 28;  DB 1;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT  10
Y05193
ID   Y05193 standard; Protein; 443 AA.
AC   Y05193;
DT   16-JUN-1999  (first entry)
DE   S. aureus priA protein sequence fragment.
KW   PriA; primosome protein  immunogenic response; T-cell immune response;
KW   respiratory tract disease; cardiac disease; gastrointestinal disease;
KW   central nervous system; kidney and urinary tract; joint disease;
KW   infection; therapy.
OS   Staphylococcus aureus.
PN   WO9910012-A1.
PD   04-MAR-1999.
PF   21-AUG-1998; U17324.
PR   25-AUG-1997; US-056939.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Warren RL;
DR   WPI; 99-204543/17.
DR   N-PSDB; X28272.
PT   Novel Staphylococcus aureus priA gene  - useful for generating
PT   immune responses in animals
PS   Example 3; Page 38; 54pp; English.
CC   This sequence is the Staphylococcus aureus priA protein of the
CC   invention. The PriA sequence (priA is a member of the primosome family)
CC   can be used to diagnose a disease related to priA expression or activity,
CC   to identify compounds which interact or inhibit priA activity, to
CC   generate an immunogenic response in a mammal so as to produce an antibody
CC   and/or T-cell immune response to protect the animal from disease.
CC   Especially mentioned is the use of the priA sequences to screen for
CC   agonists and antagonists for the treatment of e.g. upper and lower
CC   respiratory tract diseases, cardiac, gastrointestinal, central nervous
CC   system, eye, kidney and urinary tract, skin, bone and joint diseases,
CC   and bacterial (especially Staphylococcus aureus) infection.
SQ   Sequence   443 AA;

  Query Match             75.7%;  Score 28;  DB 1;  Length 443;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||:|::
Db     395 NEFRFQI 401


RESULT  11
W14849
ID   W14849 standard; Protein; 624 AA.
AC   W14849;
DT   03-JUN-1997  (first entry)
DE   Enzyme which decomposes raw starch.
KW   Enzyme; raw starch; E. coli; food; fibre; pharmaceutical.
OS   Zoogloea ramigera.
FH   Key             Location/Qualifiers
FT   peptide         1. .32
FT                   /note= "Signal peptide"
FT   protein         33. .624
FT                   /note= "Mature enzyme"
FT   region          42. .350
FT                   /note= "Active fragment"
PN   J09037789-A.
PD   10-FEB-1997.
PF   31-JUL-1995; 215457.
PR   31-JUL-1995; JP-215457.
PA   (AMAN ) AMANO PHARM KK.
DR   WPI; 97-173046/16.
```

DR N-PSDB; T64492.
PT A gene coding for an enzyme which decomposes raw starch - useful in
PT the food, fibre or pharmaceutical industry
PS Claim 1; Page 7-10; 10pp; Japanese.
CC This sequence represents an enzyme which decomposes raw starch. The
CC DNA encoding this sequence may be used to transform E. coli in the
CC production of the raw starch decomposing enzyme. The enzyme can be
CC used in various fields such as the food, fibre or pharmaceutical
CC industries, etc.
SQ Sequence 624 AA;

Query Match 75.7%; Score 28; DB 1; Length 624;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQF 5
|||||
Db 267 NEFQF 271

RESULT 12
Y05192
ID Y05192 standard; Protein; 780 AA.
AC Y05192;
DT 16-JUN-1999 (first entry)
DE S. aureus priA protein sequence.
KW PriA; primosome protein; immunogenic response; T-cell immune response;
KW respiratory tract disease; cardiac disease; gastrointestinal disease;
KW central nervous system; kidney and urinary tract; joint disease;
KW infection; therapy.
OS Staphylococcus aureus.
PN WO9910012-A1.
PD 04-MAR-1999.
PF 21-AUG-1998; U17324.
PR 25-AUG-1997; US-056939.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Warren RL;
DR WPI; 99-204543/17.
DR N-PSDB; X28269.
PT Novel Staphylococcus aureus priA gene - useful for generating
PT immune responses in animals
PS Claim 12; Page 13-14; 54pp; English.
CC This sequence is the Staphylococcus aureus priA protein of the
CC invention. The PriA sequence (priA is a member of the primosome family)
CC can be used to diagnose a disease related to priA expression or activity,
CC to identify compounds which interact or inhibit priA activity, to
CC generate an immunogenic response in a mammal so as to produce an antibody
CC and/or T-cell immune response to protect the animal from disease.
CC Especially mentioned is the use of the priA sequences to screen for
CC agonists and antagonists for the treatment of e.g. upper and lower
CC respiratory tract diseases, cardiac, gastrointestinal, central nervous
CC system, eye, kidney and urinary tract, skin, bone and joint diseases,
CC and bacterial (especially Staphylococcus aureus) infection.
SQ Sequence 780 AA;

Query Match 75.7%; Score 28; DB 1; Length 780;
Best Local Similarity 57.1%; Pred. No. 2.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
|||:|::
Db 732 NEFRFQI 738

RESULT 13
R96039
ID R96039 standard; Protein; 872 AA.
AC R96039;
DT 24-FEB-1997 (first entry)
DE 95 kD protein.
KW Tetrahymena; 80 kD protein; 95 kD protein; telomerase; telomeric repeat;
KW RNA coupled protein; chromosome; telomere repeat synthesis; antibody;
KW immortalised cell; cancer; eukaryotic microbe; inhibitor; gene therapy;
KW fungal infection; therapy; diagnosis; protozoan infection; somatic cell;
KW p80; p95.
OS Tetrahymena thermophila.
PN WO9619580-A2.
PD 27-JUN-1996.
PF 18-DEC-1995; U16531.
PR 19-DEC-1994; US-359125.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Autexier C, Collins K, Greider C, Hemish JM, Kobayashi R;
PI Yang XH;
DR WPI; 96-309594/31.
DR N-PSDB; T30868.
PT Telomerase protein and related DNA, antibodies, transgenic cells,
PT etc. - for diagnosis and treatment of cancer and infection by
PT eukaryotic microbes, also new telomerase inhibitors
PS Claim 10; Fig 4; 56pp; English.
CC R96038 and R96039 represent Tetrahymena 80 kD and 95 kD proteins (p80 and
CC p95), respectively. These proteins are components of telomerase. The
CC proteins act, when coupled to RNA, to add telomeric repeats (of the
CC sequence TTGGGG) to stabilise chromosomes. The RNA component provides
CC the template for the telomere repeat synthesis. Antibodies against these
CC sequences, can be used as immunoassay reagents for the detection of the
CC proteins. The antibodies can also be used to identify immortalised
CC cells, or predisposition to immortalisation, particularly cancer, or to
CC diagnose a disease caused by a eukaryotic microbe. Inhibitors of these
CC proteins (and these proteins themselves) can be used for therapy or
CC diagnosis. The inhibitors can be used to treat infection by fungi and
CC protozoa. As somatic cells do not generally require telomerase, these
CC inhibitors should have little or no toxicity to the host. The DNA
CC encoding these sequences are used to produce the recombinant protein, or
CC to isolate similar genes from other organisms, while transformed cells
CC can be used in gene therapy.
SQ Sequence 872 AA;

Query Match 75.7%; Score 28; DB 1; Length 872;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQF 5
|||||
Db 611 NEFQF 615

RESULT 14
Y12640
ID Y12640 standard; Protein; 76 AA.
AC Y12640;
DT 22-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO: 305 from WO 9906553.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; antiinflammatory; tumour inhibition; antitumour.
OS Homo sapiens.
PN WO9906553-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1237.
PR 01-AUG-1997; US-905051.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153783/13.
DR N-PSDB; X41498.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from umbilical cord, lymph ganglia,
PT lymphocytes and placental tissue
PS Claim 34; Page 393; 411pp; English.
CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12521 to Y12668,

CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, antiinflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 76 AA;

Query Match 73.0%; Score 27; DB 1; Length 76;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
|||.|::
Db 9 NEFAFKI 15

RESULT 15
W62301
ID W62301 standard; Protein; 539 AA.
AC W62301;
DT 28-SEP-1998 (first entry)
DE Spinach D1 protease.
KW Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;
KW detection; identification.
OS Spinacia oleracea.
PN WO9824934-A2.
PD 11-JUN-1998.
PF 03-DEC-1997; U21964.
PR 05-DEC-1996; US-759581.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI Chisholm DA, Diner BA, Donaldson GK, Hershey HP,
PI Jordan DB, Tang XS, Trost JT, Wang S, Warren PV;
DR WPI; 98-333349/29.
DR N-PSDB; V39822.
PT New isolated plant DI protease(s) - used for identifying herbicidal
PT agents which target the DI protease by inhibiting the enzyme
PT activity
PS Claim 23; Page 56-58; 73pp; English.
CC A method has been developed for identifying a herbicidal agent which
CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease
CC enzyme in a sample suspected of containing a herbicidal agent with a
CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)
CC detecting and quantifying the enzyme product of step (a). Also described
CC in the present invention are: (1) a method for detecting in vivo a
CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a
CC reaction mixture comprising: (i) a wild type cell comprising an active D1
CC protease enzyme capable of processing a D1 pre-protein; and a Photosystem
CC II core complex capable of variable chlorophyll fluorescence; (ii) a
CC suspected herbicidal agent which inhibits D1 protease; and (iii) suitable
CC growth medium; (b) illuminating the reaction mixture at illumination
CC conditions to permit D1 turnover; and (c) measuring the variable
CC chlorophyll fluorescence produced in step (b), where the level of the
CC variable chlorophyll fluorescence correlates with herbicidal activity of
CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme
CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or
CC (e) Synechocystis; (see V39818 to V39822, which encode W62297 to W62301
CC respectively). Native or recombinant D1 proteases can be used to develop
CC assays for the detection of herbicidal compositions capable of
CC inhibiting D1.
SQ Sequence 539 AA;

Query Match 73.0%; Score 27; DB 1; Length 539;
Best Local Similarity 71.4%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
| | |||
Db 28 NSFNFRV 34

Search completed: June 22, 2000, 19:15:33
Job time: 5497 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 17:59:09 ; Search time 32.22 Seconds
(without alignments)
3.137 Million cell updates/sec

Title: US-09-362-485-27
Perfect score: 37
Sequence: 1 NEFQFRV 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	28	75.7	152	2	US-08-606-143-4	Sequence 4, Appli
2	28	75.7	409	1	US-08-331-394-2	Sequence 2, Appli
3	28	75.7	409	1	US-08-250-858-2	Sequence 2, Appli
4	28	75.7	409	1	US-08-446-915-2	Sequence 2, Appli
5	28	75.7	409	2	US-08-744-139-2	Sequence 2, Appli
6	28	75.7	409	4	PCT-US95-06639-2	Sequence 2, Appli
7	27	73.0	359	2	US-09-092-770-6	Sequence 6, Appli
8	27	73.0	403	2	US-09-092-770-4	Sequence 4, Appli
9	27	73.0	404	2	US-09-092-770-3	Sequence 3, Appli
10	27	73.0	539	2	US-08-759-581B-20	Sequence 20, Appl
11	26	70.3	184	2	US-08-871-924A-4	Sequence 4, Appli
12	26	70.3	403	2	US-08-812-412-2	Sequence 2, Appli
13	26	70.3	802	1	US-07-862-021B-12	Sequence 12, Appl
14	26	70.3	802	1	US-08-313-288B-12	Sequence 12, Appl
15	26	70.3	802	4	PCT-US93-03164-12	Sequence 12, Appl
16	25	67.6	107	1	US-08-308-086-10	Sequence 10, Appl
17	25	67.6	108	2	US-08-479-078-14	Sequence 14, Appl
18	25	67.6	141	1	US-08-284-393B-10	Sequence 10, Appl
19	25	67.6	141	1	US-08-259-696B-11	Sequence 11, Appl
20	25	67.6	141	2	US-08-902-513-11	Sequence 11, Appl
21	25	67.6	141	4	PCT-US95-08950-10	Sequence 10, Appl
22	25	67.6	152	5	5187153-4	Patent No. 5187153
23	25	67.6	206	2	US-08-531-525-50	Sequence 50, Appl
24	25	67.6	206	2	US-08-718-270A-50	Sequence 50, Appl
25	25	67.6	219	2	US-08-902-516-2	Sequence 2, Appli
26	25	67.6	362	1	US-08-255-670A-4	Sequence 4, Appli
27	25	67.6	400	1	US-08-447-500-6	Sequence 6, Appli
28	25	67.6	400	1	US-08-454-097-6	Sequence 6, Appli
29	25	67.6	400	1	US-08-453-866-6	Sequence 6, Appli
30	25	67.6	415	2	US-08-860-882A-57	Sequence 57, Appl
31	25	67.6	441	2	US-08-169-948B-10	Sequence 10, Appl
32	25	67.6	441	2	US-08-448-873-10	Sequence 10, Appl
33	25	67.6	510	1	US-08-255-670A-2	Sequence 2, Appli
34	25	67.6	513	2	US-08-676-166A-4	Sequence 4, Appli
35	25	67.6	513	2	US-08-676-166A-5	Sequence 5, Appli
36	25	67.6	521	2	US-08-737-825-3	Sequence 3, Appli
37	25	67.6	525	5	5183745-4	Patent No. 5183745
38	25	67.6	590	2	US-08-756-317-12	Sequence 12, Appl
39	25	67.6	1018	1	US-08-408-093-6	Sequence 6, Appli
40	25	67.6	1018	1	US-08-408-420A-6	Sequence 6, Appli
41	25	67.6	1018	1	US-08-714-901-6	Sequence 6, Appli
42	25	67.6	1018	1	US-08-452-052-2	Sequence 2, Appli
43	25	67.6	1018	3	US-08-040-741-6	Sequence 6, Appli
44	25	67.6	1073	1	US-07-695-564-1	Sequence 1, Appli
45	25	67.6	1073	1	US-08-241-387-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-606-143-4
; Sequence 4, Application US/08606143
; Patent No. 5856155
; GENERAL INFORMATION:
;   APPLICANT: Li, Min
;   TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR
;   TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
;   TITLE OF INVENTION: COMPOUNDS
;   NUMBER OF SEQUENCES: 45
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE Leydig, Voit & Mayer, Ltd.
;     STREET: Two Prudential Plaza, Suite 4900
;     CITY: Chicago
;     STATE: IL
;     COUNTRY: USA
;     ZIP: 60601
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/606,143
;     FILING DATE: 23-FEB-1996
;     CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Kilyk Jr., John
;     REGISTRATION NUMBER: 30763
;     REFERENCE/DOCKET NUMBER: 71756
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (312) 616-5600
;     TELEFAX: (312) 616-5700
;     TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 152 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-606-143-4


  Query Match 75.7%; Score 28; DB 2; Length 152;
  Best Local Similarity 85.7%; Pred. No. 24;
  Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 NEFQFRV 7
          |||| ||
Db    112 NEFQRRV 118
```

```
RESULT   2
US-08-331-394-2
; Sequence 2, Application US/08331394
; Patent No. 5670319
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Rothe, Mike
;    TITLE OF INVENTION:  Tumor Necrosis Factor
;    TITLE OF INVENTION:  Receptor-Associated Factors
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/331,394
;      FILING DATE:
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/250858
;      FILING DATE:  27-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  897P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  409 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-331-394-2


  Query Match           75.7%;  Score 28;  DB 1;  Length 409;
  Best Local Similarity  100.0%;  Pred. No. 70;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   3
US-08-250-858-2
; Sequence 2, Application US/08250858
; Patent No. 5708142
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Rothe, Mike
;    TITLE OF INVENTION:  Tumor Necrosis Factor Receptor-Associated Factors
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/250,858
;      FILING DATE:  27-May-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  897.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  409 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-250-858-2


  Query Match           75.7%;  Score 28;  DB 1;  Length 409;
  Best Local Similarity  100.0%;  Pred. No. 70;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   4
US-08-446-915-2
; Sequence 2, Application US/08446915
; Patent No. 5741667
;  GENERAL INFORMATION:
;    APPLICANT:  Goeddel, David V.
;    APPLICANT:  Rothe. Mike
;    TITLE OF INVENTION:  Tumor Necrosis Factor Receptor-Associated Factors
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,915
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/250858
;      FILING DATE:  27-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/331394
;      FILING DATE:  28-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  897P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
```

```
;      TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  409 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
US-08-446-915-2


  Query Match             75.7%;  Score 28;  DB 1;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   5
US-08-744-139-2
; Sequence 2, Application US/08744139
; Patent No. 5869612
;  GENERAL INFORMATION:
;     APPLICANT:  Goeddel, David V
;     APPLICANT:  Rothe, Mike
;     TITLE OF INVENTION:  TUMOR NECROSIS FACTOR RECPTOR-ASSOCIATED FACTORS
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/744,139
;       FILING DATE:  31-Oct-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/250858
;       FILING DATE:  05/27/1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dreger, Ginger R.
;       REGISTRATION NUMBER:  33,055
;       REFERENCE/DOCKET NUMBER:  P0897C1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-3216
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  409 amino acids
;       TYPE:  Amino Acid
;       TOPOLOGY:  Linear
US-08-744-139-2


  Query Match             75.7%;  Score 28;  DB 2;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   6
PCT-US95-06639-2
; Sequence 2, Application PC/TUS9506639
;  GENERAL INFORMATION:
;     APPLICANT:  Genentech, Inc.
;     TITLE OF INVENTION:  Tumor Necrosis Factor Receptor-Associated Factors
;     NUMBER OF SEQUENCES:  66
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genentech, Inc.
;       STREET:  460 Point San Bruno Blvd
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  patin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06639
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/250858
;       FILING DATE:  27-MAY-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/331394
;       FILING DATE:  28-OCT-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dreger, Ginger R.
;       REGISTRATION NUMBER:  33,055
;       REFERENCE/DOCKET NUMBER:  897P2PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-3216
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  409 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
PCT-US95-06639-2


  Query Match             75.7%;  Score 28;  DB 4;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   7
US-09-092-770-6
; Sequence 6, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/092,770
;  CURRENT FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Human
US-09-092-770-6
```

```
  Query Match             73.0%;  Score 27;  DB 2;  Length 359;
  Best Local Similarity   66.7%;  Pred. No. 1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EFQFRV 7
           |||:|:
Db     245 EFQYRI 250


RESULT   8
US-09-092-770-4
; Sequence 4, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/092,770
;  CURRENT FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 403
;    TYPE: PRT
;    ORGANISM: Mouse
US-09-092-770-4


  Query Match             73.0%;  Score 27;  DB 2;  Length 403;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EFQFRV 7
           |||:|:
Db     289 EFQYRI 294


RESULT   9
US-09-092-770-3
; Sequence 3, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
;  APPLICANT: Coats, Steven R.
;  APPLICANT: Bass, Michael B.
;  APPLICANT: Robinson, Murray O.
;  TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
;  FILE REFERENCE: A-524
;  CURRENT APPLICATION NUMBER: US/09/092,770
;  CURRENT FILING DATE: 1997-06-05
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;    LENGTH: 404
;    TYPE: PRT
;    ORGANISM: Human
US-09-092-770-3


  Query Match             73.0%;  Score 27;  DB 2;  Length 404;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EFQFRV 7
           |||:|:
Db     290 EFQYRI 295


RESULT  10
US-08-759-581B-20
; Sequence 20, Application US/08759581B
; Patent No. 5876945
;  GENERAL INFORMATION:
;    APPLICANT:  CHISHOLM, DEXTER A.
;    APPLICANT:  DINER, BRUCE A.
;    APPLICANT:  DONALDSON, GAIL K.
;    APPLICANT:  HERSHEY, HOWARD P.
;    APPLICANT:  JORDAN, DOUGLAS B.
;    APPLICANT:  TANG, XIAO-SONG
;    APPLICANT:  TROST, JEFFREY T.
;    APPLICANT:  WANG, SHAOJIE
;    APPLICANT:  WARREN, PATRICK V.
;    TITLE OF INVENTION:  METHODS FOR IDENTIFYINGHERBICIDAL AGENTS THAT INHIBIT D1 PR
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/759,581B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9964
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  539 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      STRAIN:  SPINACH D1 PROTEASE PROTEIN
US-08-759-581B-20


  Query Match             73.0%;  Score 27;  DB 2;  Length 539;
  Best Local Similarity   71.4%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           | | |||
Db      28 NSFNFRV 34


RESULT  11
US-08-871-924A-4
; Sequence 4, Application US/08871924A
; Patent No. 5932469
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jeffrey D
;    TITLE OF INVENTION:  No. 5932469el Antimicrobial Polypeptides and
;    TITLE OF INVENTION:  Methods of Use
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney St
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53703
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASC II/DOS
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/871,924A
;      FILING DATE:  06/10/97
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  920308.90649
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 184 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-871-924A-4

  Query Match             70.3%;  Score 26;  DB 2;  Length 184;
  Best Local Similarity   66.7%;  Pred. No. 80;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
            |:|||:
Db       2 NDFQFQ 7


RESULT  12
US-08-812-412-2
; Sequence 2, Application US/08812412
; Patent No. 5989869
;  GENERAL INFORMATION:
;    APPLICANT:  FILIPPINI, Silvia
;    APPLICANT:  LOMOVSKAYA, Natalia
;    APPLICANT:  FONSTEIN, Leonid
;    APPLICANT:  COLOMBO, Anna L.
;    APPLICANT:  HUTCHINSON, C. R.
;    TITLE OF INVENTION:  PROCESS FOR PREPARING DAUNORUBICIN AND
;    TITLE OF INVENTION:  DOXORUBICIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/812,412
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitts, Monica C.
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-7005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-812-412-2

  Query Match             70.3%;  Score 26;  DB 2;  Length 403;
  Best Local Similarity   71.4%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
            ||| ||:
Db      38 NEFFFRL 44


RESULT  13
US-07-862-021B-12
; Sequence 12, Application US/07862021B
; Patent No. 5279966
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,021B
;      FILING DATE:  19920405
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-862-021B-12

  Query Match             70.3%;  Score 25;  DB 1;  Length 802;
  Best Local Similarity   71.4%;  Pred. No. 3.9e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
            |||  ||
Db      54 NEFSLRV 60


RESULT  14
US-08-313-288B-12
; Sequence 12, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
```

```
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-313-288B-12


  Query Match           70.3%;  Score 26;  DB 1;  Length 802;
  Best Local Similarity  71.4%;  Pred. No. 3.9e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||  ||
Db      54 NEFSLRV 60


RESULT  15
PCT-US93-03164-12
; Sequence 12, Application PC/TUS9303164
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M
;    APPLICANT:  Klar, Avihu
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03164
;      FILING DATE:  19930402
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 977-9550
;      TELEFAX:  (212) 664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  802 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-03164-12


  Query Match           70.3%;  Score 26;  DB 4;  Length 802;
  Best Local Similarity  71.4%;  Pred. No. 3.9e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||  ||
Db      54 NEFSLRV 60


Search completed: June 22, 2000, 19:27:52
Job time: 5323 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 18:02:17 ; Search time 51.21 Seconds
(without alignments)
8.014 Million cell updates/sec

Title: US-09-362-485-27
Perfect score: 37
Sequence: 1 NEFQFRV 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	91.9	1742	2	S24600	projectin - fruit
2	34	91.9	6658	2	T13931	projectin - fruit
3	34	91.9	26926	1	I38344	titin, cardiac mus
4	30	81.1	294	2	T09879	myb-related protei
5	30	81.1	562	2	T20188	hypothetical prote
6	30	81.1	864	2	S60441	hypothetical prote
7	30	81.1	1132	2	A35089	myosin-binding pro
8	29	78.4	298	2	T22390	hypothetical prote
9	29	78.4	329	2	S46807	hypothetical prote
10	29	78.4	463	2	T29671	hypothetical prote
11	29	78.4	918	2	A41667	nitrate reductase
12	29	78.4	6805	2	S20901	titin - rabbit (fr
13	28	75.7	66	2	G72804	gp41.1 protein - M
14	28	75.7	120	2	T12944	hypothetical prote
15	28	75.7	272	2	T40707	hypothetical prote
16	28	75.7	355	2	C64582	phospholipase A1 p
17	28	75.7	355	2	H71930	probable phospholi
18	28	75.7	409	2	A54750	TNF receptor assoc
19	28	75.7	434	2	T26275	hypothetical prote
20	28	75.7	506	2	B58150	hypothetical prote
21	28	75.7	506	2	T38554	hypothetical prote
22	28	75.7	528	2	I48253	beta-N-acetylhexos
23	28	75.7	538	2	T03262	cytochrome P450 -
24	28	75.7	549	2	B64507	hypothetical prote
25	28	75.7	553	2	S76949	hypothetical prote
26	28	75.7	600	2	T37497	probable CTP synth
27	28	75.7	621	1	S20145	replication factor
28	28	75.7	622	2	T27155	hypothetical prote
29	28	75.7	694	2	T30725	probable abortive
30	28	75.7	872	2	S55940	telomerase compone
31	28	75.7	940	2	A40985	projectin - fruit
32	28	75.7	964	2	T15746	hypothetical prote
33	28	75.7	997	2	S73556	MG414 homolog C12_
34	28	75.7	1123	2	S36846	myosin-binding pro
35	28	75.7	1138	2	S24614	myosin-binding pro
36	28	75.7	1142	2	S36845	myosin-binding pro
37	28	75.7	6831	2	T27934	hypothetical prote
38	28	75.7	6839	2	S57242	twitchin - Caenorh
39	28	75.7	7160	2	T27935	hypothetical prote
40	27	73.0	140	2	T36136	hypothetical prote
41	27	73.0	165	1	S55918	ribosomal protein
42	27	73.0	165	1	R3RT10	ribosomal protein
43	27	73.0	165	1	I51194	ribosomal protein
44	27	73.0	208	2	A70253	antigen, P35 homol
45	27	73.0	270	2	T38218	probable abhydrola

ALIGNMENTS

RESULT 1
S24600
projectin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1998
C;Accession: S24600
R;Fyrberg, E.
submitted to the EMBL Data Library, May 1992
A;Reference number: S24600
A;Accession: S24600
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1742 <FYR>
A;Cross-references: EMBL:X66018; NID:g8377; PID:g8378
C;Genetics:
A;Gene: FlyBase:bt
A;Cross-references: FlyBase:FBgn0005666

Query Match 91.9%; Score 34; DB 2; Length 1742;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||:||||
Db 292 NEYQFRV 298

RESULT 2
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T13931
R;Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A;Title: Structure of the Drosophila projectin protein: isoforms and implication for
A;Reference number: Z17815; MUID:98300339
A;Accession: T13931
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6658 <DAL>
A;Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1
C;Genetics:
A;Gene: projectin
A;Map position: 4
A;Note: intron positions not resolved (incomplete sequence)
C;Keywords: muscle

Query Match 91.9%; Score 34; DB 2; Length 6658;
Best Local Similarity 85.7%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFQFRV 7
             ||:||||
Db    2380 NEYQFRV 2386


RESULT   3
I38344
titin, cardiac muscle - human
N;Alternate names: connectin
N;Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R;Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:96026330
A;Accession: I38344
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMB
A;Molecule type: mRNA
A;Residues: 1-26926 <LAB1>
A;Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A;Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A;Reference number: I38345; MUID:95119041
A;Accession: I38345
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1977-2014 <MUS>
A;Cross-references: EMBL:X83270; NID:g602579; PIDN:CAA58243.1; PID:g602580
A;Note: conformation and properties are reported for a synthetic peptide corresponding t
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380
A;Accession: S20898
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 13597-14200,'I',14202-14696 <LAB2>
A;Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A;Accession: S20897
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 16330-16382,'S',16384-16756,'F',16758-16860 <LAB3>
A;Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A;Accession: S20899
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 'P',22278-22431,'R',22433-22448,'G',22450-22453,'Q',22455-22480,'TR',22483-2
A;Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A;Title: Genomic organization of M line titin and its tissue-specific expression in two
A;Reference number: S63665; MUID:96177761
A;Accession: S63665
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 26729-26825 <KOL>
A;Cross-references: EMBL:X92412; NID:g1236761
R;Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A;Reference number: S37393; MUID:94008990
A;Accession: S37393
A;Molecule type: mRNA
A;Residues: 26831-26926 <GAU>
R;Improta, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A;Reference number: A66736; PDB:1TIT
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R;Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A66201; PDB:1NCT
A;Contents: annotation; conformation by (1)H-NMR, residues 'S',26059-26155
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q32
C;Function:
A;Description: structural protein forming filaments in striated muscle
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; gl
structural protein
F;24752-25008/Domain: protein kinase homology <KIN>
F;84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,40
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,1354
tatus predicted
F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experiment


  Query Match             91.9%;  Score 34;  DB 1;  Length 26926;
  Best Local Similarity   85.7%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
             ||:||||
Db   21623 NEYQFRV 21629


RESULT   4
T09879
myb-related protein A - upland cotton
C;Species: Gossypium hirsutum (upland cotton)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
C;Accession: T09879
R;Wilkins, T.A.; Lu, C.C.
submitted to the EMBL Data Library, January 1994
A;Reference number: Z16894
A;Accession: T09879
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-294 <WIL>
A;Cross-references: EMBL:L04497; NID:g437326; PID:g437327
C;Genetics:
A;Gene: MYB A
C;Superfamily: barley myb-related protein 1; myb DNA-binding repeat homology
C;Keywords: DNA binding; transcription regulation
F;62-112/Domain: myb DNA-binding repeat homology <MYB>


  Query Match             81.1%;  Score 30;  DB 2;  Length 294;
  Best Local Similarity   83.3%;  Pred. No. 19;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
             |||||:
Db    169 NEFQFK 174


RESULT   5
T20188
hypothetical protein C53D6.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20188
R;Gardner, A.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19235
A;Accession: T20188
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-562 <WIL>
```

A;Cross-references: EMBL:Z70270; PIDN:CAA94226.1; GSPDB:GN00022; CESP:C53D6.4
A;Experimental source: clone C53D6
C;Genetics:
A;Gene: CESP:C53D6.4
A;Map position: 4
A;Introns: 15/2; 60/3; 91/1; 118/2; 227/2; 294/2; 329/2; 357/2; 396/2; 470/1; 496/1

Query Match 81.1%; Score 30; DB 2; Length 562;
Best Local Similarity 83.3%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFR 6
|||:||
Db 318 NEFEFR 323

RESULT 6
S60441
hypothetical protein YGR150c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein G6642
C;Species: Saccharomyces cerevisiae
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 29-Oct-1999
C;Accession: S60441; S64459
R;Skala, J.; Nawrocki, A.; Goffeau, A.
Yeast 11, 1421-1427, 1995
A;Title: The sequence of a 27 kb segment on the right arm of chromosome VII from Sacchar
A;Reference number: S60435; MUID:96158062
A;Accession: S60441
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-864 <SKA>
A;Cross-references: EMBL:X85807; NID:g1045249; PIDN:CAA59808.1; PID:g1045256
R;Van Dyck, L.; Skala, J.; de Wergifosse, P.; Purnelle, B.; Talla, E.; Nawrocki, A.; Del
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64428
A;Accession: S64459
A;Molecule type: DNA
A;Residues: 1-864 <VAN>
A;Cross-references: EMBL:Z72935; NID:g1323253; PIDN:CAA97164.1; PID:e243709; PID:g132325
A;Experimental source: strain S288C
C;Genetics:
A;Map position: 7R

Query Match 81.1%; Score 30; DB 2; Length 864;
Best Local Similarity 71.4%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
|||:|:|
Db 618 NEFKFKV 624

RESULT 7
A35089
myosin-binding protein C - chicken
C;Species: Gallus gallus (chicken)
C;Date: 03-Aug-1990 #sequence_revision 14-Jul-1994 #text_change 16-Jul-1999
C;Accession: A48668; A35089
R;Okagaki, T.; Weber, F.E.; Fischman, D.A.; Vaughan, K.T.; Mikawa, T.; Reinach, F.C.
J. Cell Biol. 123, 619-626, 1993
A;Title: The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides i
A;Reference number: A48668; MUID:94043444
A;Accession: A48668
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1132 <OKA>
A;Cross-references: GB:U00922; NID:g439402; PID:g439403
R;Einheber, S.; Fischman, D.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 2157-2161, 1990
A;Title: Isolation and characterization of a cDNA clone encoding avian skeletal muscle d
A;Reference number: A35089; MUID:90192766
A;Accession: A35089
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 141-1132 <EIN>
A;Cross-references: GB:M31409; NID:g212658; PID:g212659
C;Superfamily: fibronectin type III repeat homology; immunoglobulin homology
C;Keywords: actin binding; duplication; muscle
F;1053-1113/Domain: immunoglobulin homology <IMM>

Query Match 81.1%; Score 30; DB 2; Length 1132;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||::|||
Db 991 NEYRFRV 997

RESULT 8
T22390
hypothetical protein F48F7.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22390
R;Coles, L.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19559
A;Accession: T22390
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-298 <WIL>
A;Cross-references: EMBL:Z69661; PIDN:CAA93492.2; GSPDB:GN00028; CESP:F48F7.2
A;Experimental source: clone F48F7
C;Genetics:
A;Gene: CESP:F48F7.2
A;Map position: 10
A;Introns: 19/3; 48/3; 82/3; 116/1; 186/3; 237/3; 265/3

Query Match 78.4%; Score 29; DB 2; Length 298;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
|||| ||
Db 16 NEFQSRV 22

RESULT 9
S46807
hypothetical protein YHR083w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Oct-1999
C;Accession: S46807; S25455
R;Favello, T.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 9205.
A;Reference number: S46794
A;Accession: S46807
A;Molecule type: DNA
A;Residues: 1-329 <FAV>
A;Cross-references: EMBL:U10556; NID:g500825; PIDN:AAB68885.1; PID:g500828; MIPS:YHR0
R;Errede, B.; Ammerer, G.
Genes Dev. 3, 1349-1361, 1989
A;Title: STE12, a protein involved in cell-type-specific transcription and signal tra
A;Reference number: A32583; MUID:90108684
A;Accession: S25455
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 142-329 <ERR>
A;Cross-references: EMBL:X16112; NID:g4555; PIDN:CAA34245.1; PID:g832921

C;Genetics:
A;Map position: 8R

Query Match 78.4%; Score 29; DB 2; Length 329;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFR 6
|:||||
Db 228 NKFQFR 233

RESULT 10
T29671
hypothetical protein C50E3.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29671
R;Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid C50E3.
A;Reference number: Z20662
A;Accession: T29671
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-463 <GEI>
A;Cross-references: EMBL:U64848; PIDN:AAB04880.1; GSPDB:GN00023; CESP:C50E3.6
A;Experimental source: strain Bristol N2; clone C50E3
C;Genetics:
A;Gene: CESP:C50E3.6
A;Map position: 5
A;Introns: 35/1; 72/3; 110/1; 201/3; 232/1; 291/2; 334/1; 383/2

Query Match 78.4%; Score 29; DB 2; Length 463;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFR 6
|||:||
Db 235 NEFKFR 240

RESULT 11
A41667
nitrate reductase (NADH) (EC 1.6.6.1) - winter squash
C;Species: Cucurbita maxima (winter squash)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 11-Jun-1999
C;Accession: A41667
R;Hyde, G.E.; Crawford, N.M.; Campbell, W.H.
J. Biol. Chem. 266, 23542-23547, 1991
A;Title: The sequence of squash NADH:nitrate reductase and its relationship to the seque

A;Reference number: A41667; MUID:92084635
A;Accession: A41667
A;Molecule type: mRNA
A;Residues: 1-918 <HYD>
A;Cross-references: GB:M33154; NID:g167498; PIDN:AAA33114.1; PID:g167499
C;Superfamily: nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 redu
C;Keywords: chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron;
F;91-483/Domain: molybdopterin-binding domain homology <PCO>
F;543-617/Domain: cytochrome b5 core homology <CB5>
F;668-918/Domain: cytochrome-b5 reductase homology <CBR>
F;195/Binding site: molybdopterin (Cys) (covalent) #status predicted
F;434/Disulfide bonds: interchain #status predicted
F;578,601/Binding site: heme iron (His) (axial ligands) #status predicted
F;733,890/Binding site: NAD(P) (Lys, Cys) #status predicted
F;773/Binding site: FAD (Tyr) #status predicted

Query Match 78.4%; Score 29; DB 2; Length 918;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFR 6
|||:||
Db 182 NEFRFR 187

RESULT 12
S20901
titin - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C;Accession: S20901; I46520
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380
A;Accession: S20901
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-6805 <LAB>
A;Cross-references: EMBL:X64696
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
R;Labeit, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U
Nature 345, 273-276, 1990
A;Title: A regular pattern of two types of 100-residue motif in the sequence of titin
A;Reference number: I46520; MUID:90238553
A;Accession: I46520
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 4235-5250 <LA2>
A;Cross-references: EMBL:X17329; NID:g1756; PIDN:CAA35207.1; PID:g930251
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;
C;Keywords: muscle

Query Match 78.4%; Score 29; DB 2; Length 6805;
Best Local Similarity 71.4%; Pred. No. 1.1e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
:|:||||
Db 326 HEYQFRV 332

RESULT 13
G72804
gp41.1 protein - Mycobacterium phage D29
C;Species: Mycobacterium phage D29
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
C;Accession: G72804
R;Ford, M.E.; Sarkis, G.J.; Belanger, A.E.; Hendrix, R.W.; Hatfull, G.F.
J. Mol. Biol. 29, 143-164, 1998
A;Title: Genome structure of mycobacteriophage D29: Implications for phage evolution.
A;Reference number: A72800; MUID:98300335
A;Accession: G72804
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-66 <FOR>
A;Cross-references: GB:AF022214; NID:g3172250; PIDN:AAC18482.1; PID:g3172289
C;Genetics:
A;Gene: 41.1

Query Match 75.7%; Score 28; DB 2; Length 66;
Best Local Similarity 83.3%; Pred. No. 10;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EFQFRV 7
||:|||
Db 10 EFEFRV 15

RESULT 14
T12944
hypothetical protein yotI - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Nov-1999
C;Accession: T12944; C69929
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 propha
A;Reference number: Z17583
A;Accession: T12944
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-120 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025658; PIDN:AAC13153.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: C69929
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-120 <KUN>
A;Cross-references: GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13878.1; PID:e1185458;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yotI
C;Superfamily: Bacillus subtilis phage SPBc2 hypothetical protein yotI

Query Match 75.7%; Score 28; DB 2; Length 120;
Best Local Similarity 57.1%; Pred. No. 19;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||| |::
Db 77 NEFNFKI 83

RESULT 15
T40707
hypothetical protein SPBC83.18c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40707
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, February 1999
A;Reference number: A21945
A;Accession: T40707
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-272 <WOO>
A;Cross-references: EMBL:AL035536; PIDN:CAB36880.1; GSPDB:GN00067; SPDB:SPBC83.18c
A;Experimental source: strain 972h-; cosmid c83
C;Genetics:
A;Gene: SPDB:SPBC83.18c
A;Map position: 2
A;Introns: 16/3

Query Match 75.7%; Score 28; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQF 5
|||||
Db 111 NEFQF 115

Search completed: June 22, 2000, 19:28:56
Job time: 5199 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:15:37 ; Search time 40.62 Seconds
(without alignments)
5.248 Million cell updates/sec

Title: US-09-362-485-27
Perfect score: 37
Sequence: 1 NEFQFRV 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	81.1	705	1	Y4SO_RHISN	P55656 rhizobium s
2	30	81.1	864	1	YG3M_YEAST	P48237 saccharomyc
3	30	81.1	1131	1	MYPF_CHICK	P16419 gallus gall
4	29	78.4	329	1	YHO3_YEAST	P14693 saccharomyc
5	29	78.4	918	1	NIA_CUCMA	P17569 cucurbita m
6	28	75.7	409	1	TRA1_MOUSE	P39428 mus musculu
7	28	75.7	506	1	YA26_SCHPO	Q09693 schizosacch
8	28	75.7	528	1	HEXA_MOUSE	P29416 mus musculu
9	28	75.7	538	1	C7C4_MAIZE	Q43257 zea mays (m
10	28	75.7	621	1	RFA1_YEAST	P22336 saccharomyc
11	28	75.7	761	1	STAT_DROME	Q24151 drosophila
12	28	75.7	872	1	TE95_TETTH	Q94819 tetrahymena
13	28	75.7	997	1	Y414_MYCPN	P75183 mycoplasma
14	28	75.7	1141	1	MYPS_HUMAN	Q00872 homo sapien
15	28	75.7	1142	1	MYPF_HUMAN	Q14324 homo sapien
16	27	73.0	165	1	RS10_HUMAN	P46783 homo sapien
17	27	73.0	165	1	RS10_RAT	P09900 rattus norv
18	27	73.0	165	1	RS10_XENLA	Q07254 xenopus lae
19	27	73.0	374	1	NIR_RHOSH	**Q53239 rhodobacter**
20	27	73.0	404	1	CGE2_HUMAN	O96020 homo sapien
21	27	73.0	404	1	CGE2_MOUSE	Q9z238 mus musculu
22	27	73.0	457	1	V51K_ACLSA	P54892 apple chlor
23	27	73.0	460	1	V51K_ACLSP	P27739 apple chlor
24	27	73.0	782	1	CPSB_BOVIN	Q10568 bos taurus
25	27	73.0	1129	1	YB95_YEAST	P38144 saccharomyc
26	26	70.3	158	1	YRKK_BACSU	P54438 bacillus su
27	26	70.3	204	1	Y981_METJA	Q58390 methanococc
28	26	70.3	271	1	YHE9_YEAST	P38722 saccharomyc
29	26	70.3	444	1	DHE4_BACTN	P94598 bacteroides
30	26	70.3	449	1	DHE4_HAEIN	P43793 haemophilus
31	26	70.3	490	1	SBP_ARATH	O23264 arabidopsis
32	26	70.3	502	1	VL1_PCPV1	Q02274 pygmy chimp
33	26	70.3	538	1	PDP_BOVIN	P35816 bos taurus
34	26	70.3	569	1	YSY1_YEAST	P24088 saccharomyc
35	26	70.3	593	1	YG28_YEAST	P53241 saccharomyc
36	26	70.3	603	1	YH18_YEAST	P38899 saccharomyc
37	26	70.3	639	1	YGFT_ECOLI	Q46820 escherichia
38	26	70.3	758	1	YJU6_YEAST	P39529 saccharomyc
39	26	70.3	803	1	FSPO_XENLA	P35447 xenopus lae
40	26	70.3	937	1	ADB_HUMAN	P21851 homo sapien
41	26	70.3	1271	1	MYPC_CHICK	Q90688 gallus gall
42	26	70.3	1374	1	YMN3_YEAST	Q03099 saccharomyc
43	26	70.3	1681	1	YE20_YEAST	P40105 saccharomyc
44	26	70.3	1758	1	YIR7_YEAST	P40434 saccharomyc
45	26	70.3	1758	1	YJW5_YEAST	P40889 saccharomyc

ALIGNMENTS

RESULT 1
Y4SO_RHISN
ID Y4SO_RHISN STANDARD; PRT; 705 AA.
AC P55656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE PROBABLE PEPTIDASE Y4SO (EC 3.4.21.-).
GN Y4SO.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
CC PROLYL OLIGOPEPTIDASE FAMILY. STRONG, TO Y4QF.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000097; AAB91854.1; -.
DR PFAM; PF00326; Peptidase_S9; 1.
DR PRINTS; PR00862; PROLIGOPTASE.
KW Hypothetical protein; Hydrolase; Serine protease; Plasmid.
FT ACT_SITE 554 554 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 676 676 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 705 AA; 79767 MW; EA4985F696A89880 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 705;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 NEFQFRV 7
          ||| |||
Db    287 NEFLFRV 293
```

RESULT 2
YG3M_YEAST
ID YG3M_YEAST STANDARD; PRT; 864 AA.
AC P48237;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 101.4 KD PROTEIN IN RPL24B-RSR1 INTERGENIC REGION.

```
GN   YGR150C OR G6642.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 96158062.
RA   Skala J., Nawrocki A., Goffeau A.;
RT   "The sequence of a 27 kb segment on the right arm of chromosome VII
RT   from Saccharomyces cerevisiae reveals MOL1, NAT2, RPL30B, RSR1, CYS4,
RT   PEM1/CHO2, NSR1 genes and ten new open reading frames.";
RL   Yeast 11:1421-1427(1995).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85807; CAA59808.1; -.
DR   EMBL; Z72935; CAA97164.1; -.
DR   PFAM; PF01535; DUF17; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   864 AA;  101422 MW;  5A773DEAA33D84FE CRC64;

  Query Match             81.1%;  Score 30;  DB 1;  Length 864;
  Best Local Similarity   71.4%;  Pred. No. 34;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||:|:|
Db     618 NEFKFKV 624


RESULT   3
MYPF_CHICK
ID   MYPF_CHICK     STANDARD;      PRT;  1131 AA.
AC   P16419; Q90606;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MYOSIN-BINDING PROTEIN C, FAST-TYPE (FAST MYBP-C) (C-PROTEIN, SKELETAL
DE   MUSCLE FAST-ISOFORM).
GN   MYBPC2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-24; 182-198 AND 1030-1044.
RC   TISSUE=SKELETAL MUSCLE;
RX   MEDLINE; 94043444.
RA   Okagaki T., Weber F.E., Fischman D.A., Vaughan K.T., Mikawa T.,
RA   Reinach F.C.;
RT   "The major myosin-binding domain of skeletal muscle MyBP-C (C
RT   protein) resides in the COOH-terminal, immunoglobulin C2 motif.";
RL   J. Cell Biol. 123:619-626(1993).
RN   [2]
RP   SEQUENCE OF 139-1131 FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RX   MEDLINE; 90192766.
RA   Einheber S., Fischman D.A.;
RT   "Isolation and characterization of a cDNA clone encoding avian
RT   skeletal muscle C-protein: an intracellular member of the
RT   immunoglobulin superfamily.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2157-2161(1990).
CC   -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
CC       CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
CC       IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
CC       ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
CC       CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00922; AAC59644.1; -.
DR   EMBL; U00923; AAC04307.1; -.
DR   EMBL; M31209; AAA49068.1; -.
DR   PIR; A35089; A35089.
DR   HSSP; P56276; 1TLK.
DR   PFAM; PF00041; fn3; 3.
DR   PFAM; PF00047; ig; 5.
DR   PRINTS; PR00014; FNTYPEIII.
KW   Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
KW   Thick filament; Repeat.
FT   INIT_MET      0      0
FT   DOMAIN       47    148       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      248    337       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      338    428       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      429    529       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      530    629       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      630    727       FIBRONECTIN TYPE-III.
FT   DOMAIN      728    828       FIBRONECTIN TYPE-III.
FT   DOMAIN      829    922       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      923   1037       FIBRONECTIN TYPE-III.
FT   DOMAIN     1038   1131       IG-LIKE C2-TYPE DOMAIN.
SQ   SEQUENCE   1131 AA;  126811 MW;  73262DC5A939D6AC CRC64;

  Query Match             81.1%;  Score 30;  DB 1;  Length 1131;
  Best Local Similarity   71.4%;  Pred. No. 46;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||::|||
Db     990 NEYRFRV 996


RESULT   4
YHO3_YEAST
ID   YHO3_YEAST     STANDARD;      PRT;   329 AA.
AC   P14693;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   HYPOTHETICAL 37.4 KD PROTEIN IN KSP1-STE12 INTERGENIC REGION.
GN   YHR083W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 94378003.
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [2]
```

```
RP   SEQUENCE OF 142-329 FROM N.A.
RX   MEDLINE; 90108684.
RA   Errede B., Ammerer G.;
RT   "STE12, a protein involved in cell-type-specific transcription and
RT   signal transduction in yeast, is part of protein-DNA complexes.";
RL   Genes Dev. 3:1349-1361(1989).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U10556; AAB68885.1; -.
DR   EMBL; X16112; CAA34245.1; -.
DR   PIR; S25455; S25455.
DR   PIR; S46807; S46807.
KW   Hypothetical protein.
SQ   SEQUENCE   329 AA;  37404 MW;  4DCA565B56F35E4D CRC64;


  Query Match             78.4%;  Score 29;  DB 1;  Length 329;
  Best Local Similarity   83.3%;  Pred. No. 20;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           |:||||
Db     228 NKFQFR 233


RESULT   5
NIA_CUCMA
ID   NIA_CUCMA      STANDARD;      PRT;   918 AA.
AC   P17569;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   NITRATE REDUCTASE (EC 1.6.6.1) (NR).
OS   Cucurbita maxima (Pumpkin) (Winter squash).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
OC   Cucurbita.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Crawford N.M., Campbell W.H., Davis R.;
RT   "Nitrate reductase from squash: cDNA cloning and nitrate regulation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:8073-8076(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEEDLING;
RX   MEDLINE; 92084635.
RA   Hyde G.E., Crawford N.M., Campbell W.H.;
RT   "The sequence of squash NADH:nitrate reductase and its relationship
RT   to the sequences of other flavoprotein oxidoreductases. A family of
RT   flavoprotein pyridine nucleotide cytochrome reductases.";
RL   J. Biol. Chem. 266:23542-23547(1991).
CC   -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC       STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC   -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
CC   -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
CC       HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
CC       GROUP IS CALLED CYTOCHROME B-557.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- INDUCTION: BY NITRATE.
CC   -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC       N-TERMINAL DOMAIN.
CC   -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC   -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC       C-TERMINAL DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M33154; AAA33114.1; -.
DR   PIR; A41667; A41667.
DR   HSSP; P17571; 2CND.
DR   PRINTS; PR00363; CYTOCHROMEB5.
DR   PRINTS; PR00371; FPNCR.
DR   PRINTS; PR00406; CYTB5RDTASE.
DR   PRINTS; PR00407; EUMOPTERIN.
DR   PROSITE; PS00191; CYTOCHROME_B5; 1.
DR   PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
DR   PFAM; PF00173; heme_1; 1.
DR   PFAM; PF00174; oxidored_molyb; 1.
DR   PFAM; PF00175; oxidored_fad; 1.
DR   PFAM; PF00970; Cyt_reductase; 1.
KW   Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
KW   Nitrate assimilation.
FT   METAL       195    195       MOLYBDENUM-PTERIN (POTENTIAL).
FT   METAL       249    249       MOLYBDENUM-PTERIN (POTENTIAL).
FT   DISULFID    434    434       INTERCHAIN (POTENTIAL).
FT   BINDING     578    578       HEME LIGAND (BY SIMILARITY).
FT   BINDING     601    601       HEME LIGAND (BY SIMILARITY).
SQ   SEQUENCE   918 AA;  103383 MW;  A2CAFDDADEA1B2D1 CRC64;


  Query Match             78.4%;  Score 29;  DB 1;  Length 918;
  Best Local Similarity   83.3%;  Pred. No. 61;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           |||:||
Db     182 NEFRFR 187


RESULT   6
TRA1_MOUSE
ID   TRA1_MOUSE     STANDARD;      PRT;   409 AA.
AC   P39428;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   TNF RECEPTOR ASSOCIATED FACTOR 1 (TRAF1).
GN   TRAF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 123-135 AND 390-402.
RX   MEDLINE; 94349371.
RA   Rothe M., Wong S.C., Henzel W.J., Goeddel D.V.;
RT   "A novel family of putative signal transducers associated with the
RT   cytoplasmic domain of the 75 kDa tumor necrosis factor receptor.";
RL   Cell 78:681-692(1994).
CC   -!- FUNCTION: SIGNAL TRANSDUCER ASSOCIATED WITH THE CYTOPLASMIC DOMAIN
CC       OF THE 75 KD TUMOR NECROSIS FACTOR RECEPTOR (TNF-R2).
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER OF TRAF1 AND TRAF2. THIS
CC       HETEROCOMPLEX CAN BIND TO THE N-TERMINAL OF INHIBITOR OF APOPTOSIS
CC       PROTEINS 1 AND 2 (IAPS), TO RECRUIT THEM TO THE TUMOR NECROSIS
CC       FACTOR RECEPTOR 2 (TNFR2).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: CONTAINS A MATH/TRAF DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L35302; AAC37663.1; -.
DR   MGD; MGI:101836; TRAF1.
DR   PFAM; PF00917; MATH; 1.
KW   Coiled coil.
FT   DOMAIN      167    256       COILED COIL (POTENTIAL).
FT   DOMAIN      262    347       MATH/TRAF.
SQ   SEQUENCE   409 AA;  45464 MW;  EBA7FEE5639FEDDF CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db      10 NEFQF 14


RESULT   7
YA26_SCHPO
ID   YA26_SCHPO     STANDARD;      PRT;   506 AA.
AC   Q09693;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL DNA POLYMERASE BETA-LIKE PROTEIN C2F7.06C.
GN   SPAC2F7.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z50142; CAA90493.1; -.
DR   HSSP; P06766; 1BNP.
DR   PRINTS; PR00869; DNAPOLX.
DR   PROSITE; PS00522; DNA_POLYMERASE_X; 1.
DR   PFAM; PF00966; DNA_polymeraseX; 1.
KW   Hypothetical protein; Transferase.
SQ   SEQUENCE   506 AA;  57396 MW;  F9A51ED54D428134 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 506;
  Best Local Similarity   83.3%;  Pred. No. 53;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EFQFRV 7
           |||||:
Db     382 EFQFRL 387


RESULT   8
HEXA_MOUSE
ID   HEXA_MOUSE     STANDARD;      PRT;   528 AA.
AC   P29416; Q64246;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   BETA-HEXOSAMINIDASE ALPHA CHAIN PRECURSOR (EC 3.2.1.52) (N-ACETYL-
DE   BETA-GLUCOSAMINIDASE) (BETA-N-ACETYLHEXOSAMINIDASE) (HEXOSAMINIDASE
DE   A).
GN   HEXA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE; 92344643.
RA   Beccari T., Hoade J., Orlacchio A., Stirling J.L.;
RT   "Cloning and sequence analysis of a cDNA encoding the alpha-subunit
RT   of mouse beta-N-acetylhexosaminidase and comparison with the human
RT   enzyme.";
RL   Biochem. J. 285:593-596(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95135330.
RA   Bianconi S., Beccari T., Stirling J.L., Sheardown S., Orlacchio A.;
RT   "Organization of the gene for the alpha-subunit of mouse beta-N-
RT   acetylhexosaminidase (HEXa).";
RL   Biochem. Mol. Biol. Int. 34:579-586(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBA; TISSUE=LIVER;
RX   MEDLINE; 95048337.
RA   Yamanaka S., Johnson O.N., Norflus F., Boles D.J., Proia R.L.;
RT   "Structure and expression of the mouse beta-hexosaminidase genes, Hexa
RT   and Hexb.";
RL   Genomics 21:588-596(1994).
CC   -!- FUNCTION: BETA-HEXOSAMINIDASE A IS RESPONSIBLE FOR THE DEGRADATION
CC       OF GM2 GANGLIOSIDES, AND A VARIETY OF OTHER MOLECULES CONTAINING
CC       TERMINAL N-ACETYL HEXOSAMINES, IN THE BRAIN AND OTHER TISSUES.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING N-ACETYL-
CC       D-HEXOSAMINE RESIDUES IN N-ACETYL-BETA-D-HEXOSAMINIDES.
CC   -!- SUBUNIT: THERE ARE 3 FORMS OF BETA-HEXOSAMINIDASE: HEXOSAMINIDASE
CC       A IS A TRIMER COMPOSED OF ONE ALPHA CHAIN, ONE BETA-A CHAIN & ONE
CC       BETA-B CHAIN. HEXOSAMINIDASE B IS A TETRAMER OF TWO BETA-A AND TWO
CC       BETA-B CHAINS. HEXOSAMINIDASE S IS AN HOMODIMER OF TWO ALPHA
CC       CHAINS. THE TWO BETA CHAINS ARE DERIVED FROM THE CLEAVAGE OF A
CC       PRECURSOR CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES. STRONG,
CC       TO BETA CHAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X64331; CAA45615.1; -.
DR   EMBL; U05837; AAC53246.1; -.
DR   EMBL; U05824; AAC53246.1; JOINED.
DR   EMBL; U05825; AAC53246.1; JOINED.
DR   EMBL; U05826; AAC53246.1; JOINED.
DR   EMBL; U05827; AAC53246.1; JOINED.
DR   EMBL; U05828; AAC53246.1; JOINED.
DR   EMBL; U05829; AAC53246.1; JOINED.
DR   EMBL; U05830; AAC53246.1; JOINED.
DR   EMBL; U05831; AAC53246.1; JOINED.
DR   EMBL; U05832; AAC53246.1; JOINED.
DR   EMBL; U05833; AAC53246.1; JOINED.
DR   EMBL; U05834; AAC53246.1; JOINED.
DR   EMBL; U05835; AAC53246.1; JOINED.
DR   EMBL; U05836; AAC53246.1; JOINED.
DR   EMBL; U07721; AAA18777.1; -.
DR   EMBL; U07709; AAA18777.1; JOINED.
DR   EMBL; U07710; AAA18777.1; JOINED.
DR   EMBL; U07711; AAA18777.1; JOINED.
```

```
DR   EMBL; U07712; AAA18777.1; JOINED.
DR   EMBL; U07713; AAA18777.1; JOINED.
DR   EMBL; U07714; AAA18777.1; JOINED.
DR   EMBL; U07715; AAA18777.1; JOINED.
DR   EMBL; U07716; AAA18777.1; JOINED.
DR   EMBL; U07717; AAA18777.1; JOINED.
DR   EMBL; U07718; AAA18777.1; JOINED.
DR   EMBL; U07719; AAA18777.1; JOINED.
DR   EMBL; U07720; AAA18777.1; JOINED.
DR   EMBL; U07631; AAA18775.1; -.
DR   PIR; S24323; S24323.
DR   HSSP; P06865; 1QBC.
DR   MGD; MGI:96073; HEXA.
DR   PFAM; PF00728; Glyco_hydro_20; 1.
DR   PRINTS; PR00738; GLHYDRLASE20.
KW   Hydrolase; Glycosidase; Lysosome; Signal; Zymogen; Glycoprotein.
FT   SIGNAL       1     22       BY SIMILARITY.
FT   PROPEP      23    108       BY SIMILARITY.
FT   CHAIN      109    528       BETA-HEXOSAMINIDASE ALPHA CHAIN.
FT   CARBOHYD   115    115       POTENTIAL.
FT   CARBOHYD   157    157       POTENTIAL.
FT   CARBOHYD   295    295       POTENTIAL.
FT   CARBOHYD   487    487       POTENTIAL.
FT   ACT_SITE   323    323       CATALYTIC ACID (BINDS TO THE GLYCOSIDIC
FT                               LINKAGE) (BY SIMILARITY).
FT   CONFLICT    56     56       G -> A (IN REF. 3).
SQ   SEQUENCE   528 AA;  60599 MW;  5C723B82DAEC3B42 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 528;
  Best Local Similarity   83.3%;  Pred. No. 56;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           | ||||
Db      42 NNFQFR 47


RESULT   9
C7C4_MAIZE
ID   C7C4_MAIZE     STANDARD;      PRT;   538 AA.
AC   Q43257; O04990;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CYTOCHROME P450 71C4 (EC 1.14.-.-).
GN   CYP71C4.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC   Poaceae; Zea.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CI31A;
RX   MEDLINE; 95124291.
RA   Frey M., Kliem R., Saedler H., Gierl A.;
RT   "Expression of a cytochrome P450 gene family in maize.";
RL   Mol. Gen. Genet. 246:100-109(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CI31A;
RA   Gierl A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X81831; CAA57425.1; -.
DR   EMBL; Y11368; CAA72196.1; -.
DR   PROSITE; PS00086; CYTOCHRC_E_P450; 1.
DR   PFAM; PF00067; p450; 1.
KW   Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT   TRANSMEM     22     42       POTENTIAL.
FT   BINDING     481    481       HEME (BY SIMILARITY).
SQ   SEQUENCE   538 AA;  60047 MW;  D2F8CC0E62891D46 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 538;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db     467 NEFQF 471


RESULT  10
RFA1_YEAST
ID   RFA1_YEAST     STANDARD;      PRT;   621 AA.
AC   P22336; P38906;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   REPLICATION FACTOR-A PROTEIN 1 (RF-A) (SINGLE-STRANDED DNA-BINDING
DE   PROTEIN) (DNA BINDING PROTEIN BUF2) (REPLICATION PROTEIN A 69 KD DNA-
DE   BINDING SUBUNIT).
GN   RFA1 OR BUF2 OR RPA1 OR YAR007C OR FUN3.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 111-129 AND 495-503.
RC   STRAIN=W303-1A;
RX   MEDLINE; 91357474.
RA   Brill S.J., Stillman B.;
RT   "Replication factor-A from Saccharomyces cerevisiae is encoded by
RT   three essential genes coordinately expressed at S phase.";
RL   Genes Dev. 5:1589-1600(1991).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 90291999.
RA   Heyer W.-D., Rao M.R., Erdile L.F., Kelley T.J., Kolodner R.D.;
RT   "An essential Saccharomyces cerevisiae single-stranded DNA binding
RT   protein is homologous to the large subunit of human RP-A.";
RL   EMBO J. 9:2321-2329(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93361008.
RA   Luche R.M., Smart W.C., Marion T., Tillman M., Sumrada R.A.,
RA   Cooper T.G.;
RT   "Saccharomyces cerevisiae BUF protein binds to sequences
RT   participating in DNA replication in addition to those mediating
RT   transcriptional repression (URS1) and activation.";
RL   Mol. Cell. Biol. 13:5749-5761(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 95028152.
RA   Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
RA   Delaney S., Ouellette B.F.F., Barton A.B., Kaback D.B., Bussey H.;
RT   "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
RT   the 42 kbp SPO7-CENI-CDC15 region.";
RL   Yeast 10:535-541(1994).
CC   -!- FUNCTION: BINDS TO SINGLE-STRANDED SEQUENCES PARTICIPATING IN DNA
CC       REPLICATION IN ADDITION TO THOSE MEDIATING TRANSCRIPTIONAL
CC       REPRESSION (URS1) AND ACTIVATION (CAR1) STIMULATES THE ACTIVITY
CC       OF A COGNATE STRAND EXCHANGE PROTEIN (SEP1). IT COOPERATES WITH
CC       T-AG AND DNA TOPOISOMERASE I TO UNWIND TEMPLATE DNA CONTAINING THE
CC       SIMIAN VIRUS 40 ORIGIN OF DNA REPLICATION.
CC   -!- SUBUNIT: HETEROTRIMER OF 69, 36, AND 13 KD CHAINS. THE DNA-BINDING
```

```
CC       ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 69 KD SUBUNIT.
CC    -!- SUBCELLULAR LOCATION: NUCLEAR.
CC    -!- SIMILARITY: TO OTHER SPECIES RFA1/RPA1.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; X59748; CAA42420.1; -.
DR    EMBL; M60262; AAA34994.1; -.
DR    EMBL; S64901; AAB27889.1; -.
DR    EMBL; L22015; AAC04960.1; -.
DR    PIR; A37281; A37281.
DR    PIR; S20145; S20145.
DR    PIR; S40902; S40902.
DR    HSSP; P27694; 1JMC.
DR    TRANSFAC; T01250; -.
DR    SGD; L0001620; RFA1.
KW    DNA replication; DNA-binding; Zinc-finger; Nuclear protein.
FT    MOD_RES      ?1     ?1       BLOCKED.
FT    ZN_FING     486    508       C4-TYPE (POTENTIAL).
SQ    SEQUENCE   621 AA;  70347 MW;  7EB8DFA6910EF8A8 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 621;
  Best Local Similarity   57.1%;  Pred. No. 66;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||: ||:
Db     573 NEYDFRI 579


RESULT  11
STAT_DROME
ID    STAT_DROME     STANDARD;      PRT;   761 AA.
AC    Q24151; Q24181;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    01-NOV-1997 (Rel. 35, Last annotation update)
DE    SIGNAL TRANSDUCER AND TRANSCRIPTION ACTIVATOR (MARELLE PROTEIN)
DE    (D-STAT).
GN    STAT92E OR MARE OR MRL OR STAT.
OS    Drosophila melanogaster (Fruit fly).
OC    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC    Ephydroidea; Drosophilidae; Drosophila.
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 96190808.
RA    Hou X.S., Melnick M.B., Perrimon N.;
RT    "Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes
RT    a protein similar to the mammalian STATs.";
RL    Cell 84:411-419(1996).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    TISSUE=EMBRYO;
RX    MEDLINE; 96190809.
RA    Yan R., Small S., Desplan C., Dearolf C.R., Darnell J.E. Jr.;
RT    "Identification of a Stat gene that functions in Drosophila
RT    development.";
RL    Cell 84:421-430(1996).
CC    -!- FUNCTION: SIGNAL TRANSDUCTION AND ACTIVATION OF TRANSCRIPTION.
CC        PLAYS AN IMPORTANT ROLE IN THE SEGMENTAL PATTERN FORMATION IN THE
CC        EARLY EMBRYO BY ACTIVATING SPECIFIC STRIPES OF PAIR RULE GENE
CC        EXPRESSION. THE JANUS KINASE-STAT PATHWAY IS CONNECTED TO
CC        DROSOPHILA EARLY DEVELOPMENT. MUTANTS EXHIBIT ABERRANT EXPRESSION
CC        OF THE PAIR RULE GENE EVEN-SKIPPED AT THE CELLULAR BLASTODERM
CC        STAGE, LEADING TO LARVAL SEGMENTATION DEFECTS.
CC    -!- SUBUNIT: FORMS A HOMODIMER OR A HETERODIMER WITH A RELATED FAMILY
CC        MEMBER.
CC    -!- SUBCELLULAR LOCATION: NUCLEAR; TRANSLOCATED INTO THE NUCLEUS IN
CC        RESPONSE TO PHOSPHORYLATION (BY SIMILARITY).
CC    -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE
CC        SPLICING. THE SEQUENCE SHOWN HERE IS THAT OF THE LONG FORM.
CC    -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL,
CC        PUPAL, AND ADULT STAGES, WITH SOME DECREASE IN THE LATE EMBRYONIC
CC        STAGES. THE EXPRESSION IS UNIFORM IN UNFERTILIZED OR JUST
CC        FERTILIZED EGGS, SUGGESTING MATERNALLY DEPOSITED MRNA. AT
CC        BLASTODERM STAGE, EXPRESSION PATTERN SHOWS STRIPES, THAT ARE
CC        REMINISCENT OF MANY PAIR RULE GENES PATTERN.
CC    -!- PTM: TYROSINE PHOSPHORYLATED BY HOPSCOTCH. PHOSPHORYLATION IS
CC        REQUIRED FOR DNA-BINDING ACTIVITY AND DIMERIZATION.
CC    -!- SIMILARITY: BELONGS TO THE STAT FAMILY OF TRANSCRIPTION FACTORS.
CC        IN PARTICULAR TO MAMMALIAN STAT5 AND STAT6.
CC    -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; U40070; AAC46984.1; -.
DR    EMBL; U46688; AAB02195.1; -.
DR    HSSP; P42224; 1BF5.
DR    FLYBASE; FBgn0016917; Stat92E.
DR    PROSITE; PS50001; SH2; 1.
DR    PFAM; PF01017; STAT; 1.
KW    Transcription regulation; DNA-binding; Nuclear protein; SH2 domain;
KW    Phosphorylation; Alternative splicing; Developmental protein.
FT    DOMAIN      594    658       SH2.
FT    MOD_RES     711    711       PHOSPHORYLATION (BY JAKS).
FT    VARSPLIC    699    705       MISSING (IN SHORT ISOFORM).
FT    CONFLICT    105    105       S -> T (IN REF. 2).
FT    CONFLICT    648    648       L -> H (IN REF. 2).
SQ    SEQUENCE   761 AA;  86414 MW;  BF3A622A29899161 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 761;
  Best Local Similarity   71.4%;  Pred. No. 83;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |:|| ||
Db     467 NDFQIRV 473


RESULT  12
TE95_TETTH
ID    TE95_TETTH     STANDARD;      PRT;   872 AA.
AC    Q94819;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    15-JUL-1998 (Rel. 36, Last annotation update)
DE    TELOMERASE COMPONENT P95 (EC 2.7.7.-).
OS    Tetrahymena thermophila.
OC    Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC    Tetrahymenina; Tetrahymena.
RN    [1]
RP    SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX    MEDLINE; 95292335.
RA    Collins K., Kobayashi R., Greider C.W.;
RT    "Purification of Tetrahymena telomerase and cloning of genes encoding
RT    the two protein components of the enzyme.";
RL    Cell 81:677-686(1995).
CC    -!- FUNCTION: RIBONUCLEOPROTEIN DNA POLYMERASE THAT CATALYZES THE DE
CC        NOVO SYNTHESIS OF TELOMERIC SIMPLE SEQUENCE REPEATS. P95 CONTAINS
CC        SOME OR ALL OF THE TEMPLATE-INDEPENDENT PRIMER DNA-BINDING SITE
CC        TERMED THE ANCHOR SITE.
```

```
CC   -!- SUBUNIT: TELOMERASE CONSIST OF TWO SUBUNIT, P80 AND P95 THAT FORM
CC       A 1:1:1 COMPLEX WITH THE 159 NT TELOMERASE RNA.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25642; AAC46602.1; -.
KW   Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
KW   DNA-binding.
SQ   SEQUENCE   872 AA;  102946 MW;  78AFF4CD829651FA CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 872;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db     611 NEFQF 615


RESULT  13
Y414_MYCPN
ID   Y414_MYCPN     STANDARD;      PRT;   997 AA.
AC   P75183;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HYPOTHETICAL PROTEIN MG414.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- SIMILARITY: BELONGS TO THE MG414 / MG415 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000023; AAB95878.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   997 AA;  117140 MW;  B8E903709A6FCBB6 CRC64;


  Query Match             75.7%;  Score 28;  DB 1;  Length 997;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           | ||||
Db     966 NNFQFR 971


RESULT  14
MYPS_HUMAN
ID   MYPS_HUMAN     STANDARD;      PRT;  1141 AA.
AC   Q00872; Q15497;
DT   01-FEB-1994 (Rel. 28, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MYOSIN-BINDING PROTEIN C, SLOW-TYPE (SLOW MYBP-C) (C-PROTEIN, SKELETAL
DE   MUSCLE SLOW-ISOFORM).
GN   MYBPC1 OR MYBPCS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL SKELETAL MUSCLE;
RX   MEDLINE; 93054997.
RA   Fuerst D.O., Vinkemeier U., Weber K.;
RT   "Mammalian skeletal muscle C-protein: purification from bovine muscle,
RT   binding to titin and the characterization of a full-length human
RT   cDNA.";
RL   J. Cell Sci. 102:769-778(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL SKELETAL MUSCLE;
RX   MEDLINE; 93387319.
RA   Weber F.E., Vaughan K.T., Reinach F.C., Fischman D.A.;
RT   "Complete sequence of human fast-type and slow-type muscle myosin-
RT   binding-protein C (MyBP-C). Differential expression, conserved domain
RT   structure and chromosome assignment.";
RL   Eur. J. Biochem. 216:661-669(1993).
CC   -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
CC       CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
CC       IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
CC       ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
CC       CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66276; CAA46987.1; -.
DR   EMBL; X73114; CAA51545.1; -.
DR   PIR; S24614; S24614.
DR   MIM; 160794; -.
DR   PFAM; PF00041; fn3; 3.
DR   PFAM; PF00047; ig; 2.
KW   Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
KW   Thick filament; Repeat.
FT   DOMAIN       72    144       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      251    340       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      341    431       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      432    520       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      522    619       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      620    717       FIBRONECTIN TYPE-III.
FT   DOMAIN      718    836       FIBRONECTIN TYPE-III.
FT   DOMAIN      837    931       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      932   1046       FIBRONECTIN TYPE-III.
FT   DOMAIN     1047   1141       IG-LIKE C2-TYPE DOMAIN.
FT   CONFLICT     87     88       LR -> SE (IN REF. 1).
FT   CONFLICT     93     97       KWFKG -> NGSR (IN REF. 1).
FT   CONFLICT    217    217       A -> T (IN REF. 1).
FT   CONFLICT    230    235       GITDLR -> ESPTCS (IN REF. 1).
FT   CONFLICT    244    245       MR -> SI (IN REF. 1).
FT   CONFLICT    261    261       A -> V (IN REF. 1).
FT   CONFLICT    286    286       Y -> N (IN REF. 1).
FT   CONFLICT    292    292       I -> L (IN REF. 1).
FT   CONFLICT    302    312       HKGCQRILFIN -> DTRCQSILNID (IN REF. 1).
```

```
FT   CONFLICT    336    336       F -> L (IN REF. 1).
FT   CONFLICT    353    353       A -> D (IN REF. 1).
FT   CONFLICT    371    371       N -> Q (IN REF. 1).
FT   CONFLICT    383    387       PGPKS -> LVQT (IN REF. 1).
FT   CONFLICT    412    412       E -> D (IN REF. 1).
FT   CONFLICT    491    492       AN -> DH (IN REF. 1).
FT   CONFLICT    761    779       STSAKQSDENGEAAYDLPA -> T (IN REF. 2).
FT   CONFLICT    840    843       RIPR -> HSPK (IN REF. 1).
FT   CONFLICT    855    858       EAVN -> DRVI (IN REF. 1).
FT   CONFLICT    927    927       Q -> R (IN REF. 1).
FT   CONFLICT    946    946       E -> R (IN REF. 1).
FT   CONFLICT    979    980       FT -> LR (IN REF. 1).
FT   CONFLICT    985    992       YHRTSATI -> IIEPVPH (IN REF. 1).
FT   CONFLICT   1056   1063       TYAIAGYN -> RLCHSGYM (IN REF. 1).
FT   CONFLICT   1099   1099       Q -> L (IN REF. 1).
FT   CONFLICT   1107   1107       R -> G (IN REF. 1).
SQ   SEQUENCE   1141 AA;  128294 MW;  8B3A91422CA9F2B2 CRC64;

  Query Match             75.7%;  Score 28;  DB 1;  Length 1141;
  Best Local Similarity   71.4%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
            ||: |||
Db     999 NEYYFRV 1005


RESULT  15
MYPF_HUMAN
ID   MYPF_HUMAN     STANDARD;      PRT;  1142 AA.
AC   Q14324;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MYOSIN-BINDING PROTEIN C, FAST-TYPE (FAST MYBP-C) (C-PROTEIN, SKELETAL
DE   MUSCLE FAST-ISOFORM).
GN   MYBPC2 OR MYBPCF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL SKELETAL MUSCLE;
RX   MEDLINE; 93387319.
RA   Weber F.E., Vaughan K.T., Reinach F.C., Fischman D.A.;
RT   "Complete sequence of human fast-type and slow-type muscle myosin-
RT   binding-protein C (MyBP-C). Differential expression, conserved domain
RT   structure and chromosome assignment.";
RL   Eur. J. Biochem. 216:661-669(1993).
CC   -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
CC       CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
CC       IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
CC       ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
CC       CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73113; CAA51544.1; -.
DR   MIM; 160793; -.
DR   PFAM; PF00041; fn3; 3.
DR   PFAM; PF00047; ig; 3.
DR   PRINTS; PR00014; FNTYPEIII.
KW   Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
KW   Thick filament; Repeat.
FT   DOMAIN       50    153       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      255    344       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      345    437       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      438    539       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      540    639       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      640    737       FIBRONECTIN TYPE-III.
FT   DOMAIN      738    838       FIBRONECTIN TYPE-III.
FT   DOMAIN      839    933       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      934   1048       FIBRONECTIN TYPE-III.
FT   DOMAIN     1049   1142       IG-LIKE C2-TYPE DOMAIN.
SQ   SEQUENCE   1142 AA;  128142 MW;  93461D4356E86C09 CRC64;

  Query Match             75.7%;  Score 28;  DB 1;  Length 1142;
  Best Local Similarity   71.4%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
            ||: |||
Db    1001 NEYYFRV 1007


Search completed: June 22, 2000, 19:33:55
Job time: 1098 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:58:59 ; Search time 783.2 Seconds
(without alignments)
-18.631 Million cell updates/sec

Title: US-09-362-485-28
Perfect score: 15
Sequence: 1 AACGAATTCCGGGTG 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_in1:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*
44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*
51: gb_pr5:*
52: gb_htg8:*
53: gb_htg9:*
54: gb_htg10:*
55: gb_htg11:*
56: gb_htg12:*
57: gb_htg13:*
58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		15	100.0	15	5	A87631	A87631 Sequence 28
2		15	100.0	15	5	A89756	A89756 Sequence 13
3		15	100.0	1194	2	MTU92472	U92472 Mycobacteri
4		15	100.0	1208	5	A89752	A89752 Sequence 9
5		15	100.0	1209	5	A87612	A87612 Sequence 9
6		15	100.0	1228	5	A87608	A87608 Sequence 5
7		15	100.0	1228	5	A89748	A89748 Sequence 5
8		15	100.0	1229	5	A87610	A87610 Sequence 7
9		15	100.0	1229	5	A89750	A89750 Sequence 7
10		15	100.0	1235	5	A87606	A87606 Sequence 3
11		15	100.0	1235	5	A87609	A87609 Sequence 6
12		15	100.0	1235	5	A87611	A87611 Sequence 8
13		15	100.0	1235	5	A89746	A89746 Sequence 3
14		15	100.0	1235	5	A89749	A89749 Sequence 6
15		15	100.0	1235	5	A89751	A89751 Sequence 8
16		15	100.0	1236	5	A87613	A87613 Sequence 10
17		15	100.0	1236	5	A89753	A89753 Sequence 10
18		15	100.0	1237	5	A87607	A87607 Sequence 4
19		15	100.0	1237	5	A89747	A89747 Sequence 4
20		15	100.0	56414	1	MTV002	AL008967 Mycobacte
21	c	14	93.3	14852	1	EC4HPADNA	Z37980 E.coli hpa[
22		14	93.3	66146	44	AC020262	AC020262 Drosophil
23	c	14	93.3	178328	43	AC007299	AC007299 Drosophil
24		14	93.3	185747	54	AC018978	AC018978 Homo sapi
25	c	13.4	89.3	305	13	G35956	G35956 STS h14a134
26	c	13.4	89.3	405	13	G21090	G21090 human STS W
27		13.4	89.3	567	9	HUMISGA2	M21786 Human inter
28		13.4	89.3	634	9	HUMIFN15K	M13755 Human inter
29		13.4	89.3	696	8	CNS0199Z	AL111375 Botrytis
30		13.4	89.3	1043	7	LAGLUSYN	X15578 Lupinus ang
31		13.4	89.3	1344	7	BNGSR12	Y12459 Brassica na
32		13.4	89.3	1400	7	BNGLUTS	X76736 B.napus mRN
33		13.4	89.3	1506	16	AF083498	AF083498 Peanut st
34	c	13.4	89.3	2518	40	AF019634	AF019634 Homo sapi
35		13.4	89.3	2523	34	DROHAYWIRA	L02965 Drosophila
36		13.4	89.3	2578	34	CMERCC3A	X68309 D.melanogas
37		13.4	89.3	2750	10	HUMPPA	L02867 Homo sapien
38	c	13.4	89.3	2785	10	HSU59305	U59305 Human ser-t
39		13.4	89.3	2785	12	D50656	D50656 Mus musculu
40		13.4	89.3	3302	4	AF175313	AF175313 Oncorhync
41		13.4	89.3	3673	1	RMAJ661	AJ222661 Rhizobium
42		13.4	89.3	4849	2	AFAPHB	J04223 Ralstonia p
43		13.4	89.3	4857	2	AF112870	AF112870 Pseudomon
44	c	13.4	89.3	8586	51	AF203815	AF203815 Homo sapi
45		13.4	89.3	10808	2	AE001235	AE001235 Treponema

ALIGNMENTS

```
RESULT   1
A87631
LOCUS       A87631         15 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 28 from Patent WO9836089.
ACCESSION   A87631
VERSION     A87631.1  GI:6736270
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      3 c      5 g      3 t
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 15;
  Best Local Similarity 100.0%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db      1 AACGAATTCCGGGTG 15


RESULT   2
A89756
LOCUS       A89756         15 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 13 from Patent WO9832862.
ACCESSION   A89756
VERSION     A89756.1  GI:6738289
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        4 a      3 c      5 g      3 t
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 15;
  Best Local Similarity 100.0%;  Pred. No. 4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db      1 AACGAATTCCGGGTG 15


RESULT   3
MTU92472
LOCUS       MTU92472     1194 bp    DNA             BCT       07-AUG-1998
DEFINITION  Mycobacterium tuberculosis L-alanine dehydrogenase gene, complete
            cds.
ACCESSION   U92472
VERSION     U92472.1  GI:3089350
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Andersen,A.B. and Hansen,E.B.
  TITLE     Structure and mapping of antigenic domains of protein antigen b, a
            38,000-molecular-weight protein of Mycobacterium tuberculosis
  JOURNAL   Infect. Immun. 57 (8), 2481-2488 (1989)
  MEDLINE   89307568
REFERENCE   2  (bases 1 to 1194)
  AUTHORS   Hutter,B. and Singh,M.
  TITLE     Host vector system for high-level expression and purification of
            recombinant, enzymatically active alanine dehydrogenase of
            mycobacterium tuberculosis
  JOURNAL   Gene 212 (1), 21-29 (1998)
  MEDLINE   98267225
REFERENCE   3  (bases 1 to 1194)
  AUTHORS   Singh,M. and Hutter,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1997) GBF, Mascheroder Weg 1, Braunschweig 38124,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37rv and H37ra"
                     /db_xref="taxon:1773"
     CDS             61. .1176
                     /EC_number="1.4.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="L-alanine dehydrogenase"
                     /protein_id="AAC38804.1"
                     /db_xref="GI:3089351"
                     /translation="MRVGIPTETKNNEFRVAITPAGVAELTRRGHEVLIQAGAGEGSA
                     ITDADFKAAGAQLVGTADQVWADADLLLKVKEPIAAEYGRLRHGQILFTFLHLAASRA
                     CTDALLDSGTTSIAYETVQTADGALPLLAPMSEVAGRLAAQVGAYHLMRTQGGRGVLM
                     GGVPGVEPADVVVIGAGTAGYNAARIANGMGATVTVLDINIDKLRQLDAEFCGRIHTR
                     YSSAYELEGAVKRADLVIGAVLVPGAKAPKLVSNSLVAHMKPGAVLVDIAIDQGGCFE
                     GSRPTTYDHPTFAVHDTLFYCVANMPASVPKTSTYALTNATMPYVLELADHGWRAACR
                     SNPALAKGLSTHEGALLSERVATDLGVPFTEPASVLA"
BASE COUNT      226 a    385 c    368 g    215 t
ORIGIN

  Query Match           100.0%;  Score 15;  DB 2;  Length 1194;
  Best Local Similarity 100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db     94 AACGAATTCCGGGTG 108


RESULT   4
A89752
LOCUS       A89752       1208 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9832862.
ACCESSION   A89752
VERSION     A89752.1  GI:6738286
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1208)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="unidentified"
```

```
                     /db_xref="taxon:32644"
BASE COUNT       228 a    391 c    373 g    216 t
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT   5
A87612
LOCUS       A87612       1209 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 9 from Patent WO9836089.
ACCESSION   A87612
VERSION     A87612.1  GI:6736252
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1209)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       228 a    391 c    373 g    216 t      1 others
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1209;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT   6
A87608
LOCUS       A87608       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9836089.
ACCESSION   A87608
VERSION     A87608.1  GI:6736248
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT   7
A89748
LOCUS       A89748       1228 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 5 from Patent WO9832862.
ACCESSION   A89748
VERSION     A89748.1  GI:6738282
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1228;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT   8
A87610
LOCUS       A87610       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9836089.
ACCESSION   A87610
VERSION     A87610.1  GI:6736250
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       236 a    391 c    382 g    219 t      1 others
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1229;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT   9
A89750
LOCUS       A89750       1229 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 7 from Patent WO9832862.
```

```
ACCESSION   A89750
VERSION     A89750.1  GI:6738284
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1229)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1229
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    391 c    382 g    219 t      1 others
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 1229;
  Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db     94 AACGAATTCCGGGTG 108


RESULT  10
A87606
LOCUS       A87606       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9836089.
ACCESSION   A87606
VERSION     A87606.1  GI:6736246
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db     94 AACGAATTCCGGGTG 108


RESULT  11
A87609
LOCUS       A87609       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9836089.
ACCESSION   A87609
VERSION     A87609.1  GI:6736249
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db     94 AACGAATTCCGGGTG 108


RESULT  12
A87611
LOCUS       A87611       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9836089.
ACCESSION   A87611
VERSION     A87611.1  GI:6736251
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     TEST KIT FOR TUBERCULOSIS DIAGNOSIS OR THE LIKE
  JOURNAL   Patent: WO 9836089-A 20-AUG-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN

  Query Match           100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||||||||||||||
Db     94 AACGAATTCCGGGTG 108


RESULT  13
A89746
LOCUS       A89746       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9832862.
ACCESSION   A89746
VERSION     A89746.1  GI:6736280
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    382 g    220 t      3 others
ORIGIN
```

```
  Query Match             100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT  14
A89749
LOCUS       A89749       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent WO9832862.
ACCESSION   A89749
VERSION     A89749.1  GI:6738283
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      235 a    395 c    384 g    220 t      1 others
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108


RESULT  15
A89751
LOCUS       A89751       1235 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 8 from Patent WO9832862.
ACCESSION   A89751
VERSION     A89751.1  GI:6738285
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1235)
  AUTHORS   Flohe,L. and Singh,M.
  TITLE     L-ALANINE DEHYDROGENASE OF MYCOBACTERIUM MARINUM
  JOURNAL   Patent: WO 9832862-A 30-JUL-1998;
            FLOHE LEOPOLD (DE); SINGH MAHAVIR (DE)
FEATURES             Location/Qualifiers
     source          1. .1235
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      236 a    394 c    385 g    220 t
ORIGIN


  Query Match             100.0%;  Score 15;  DB 5;  Length 1235;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||||||||
Db      94 AACGAATTCCGGGTG 108
```

Search completed: June 22, 2000, 17:59:01
Job time: 10810 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 18:04:18 ; Search time 54.21 Seconds
(without alignments)
8.953 Million cell updates/sec

Title: US-09-362-485-27
Perfect score: 37
Sequence: 1 NEFQFRV 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	91.9	1742	5	Q24463	Q24463 drosophila
2	34	91.9	6658	5	O76281	O76281 drosophila
3	34	91.9	26926	4	Q10466	Q10466 homo sapien
4	30	81.1	294	10	Q43436	Q43436 gossypium h
5	30	81.1	562	5	Q18810	Q18810 caenorhabdi
6	29	78.4	463	5	Q22925	Q22925 caenorhabdi
7	29	78.4	6875	6	Q28733	Q28733 oryctolagus
8	28	75.7	66	9	O64232	O64232 mycobacteri
9	28	75.7	120	9	O64191	O64191 bacteriopha
10	28	75.7	186	8	e30705	e30705 bacillus su
11	28	75.7	272	5	O94701	O94701 schizosacch
12	28	75.7	355	2	O25241	O25241 helicobacte
13	28	75.7	355	2	Q9ZLX5	Q9zlx5 helicobacte
14	28	75.7	434	5	Q23217	Q23217 caenorhabdi
15	28	75.7	488	5	Q25376	Q25376 loligo opal
16	28	75.7	515	5	Q16968	Q16968 aplysia cal
17	28	75.7	528	11	Q64246	Q64246 mus musculu
18	28	75.7	549	1	Q59054	Q59054 methanococc
19	28	75.7	553	2	P74741	P74741 synechocyst
20	28	75.7	600	3	O42644	O42644 schizosacch
21	28	75.7	622	5	Q9XWK4	Q9xwk4 caenorhabdi
22	28	75.7	694	12	Q98290	Q98290 molluscum c
23	28	75.7	798	4	Q9Y5E9	Q9y5e9 homo sapien
24	28	75.7	964	5	Q18382	Q18382 caenorhabdi
25	28	75.7	1402	5	O96668	O96668 dictyosteli
26	28	75.7	6048	5	Q23020	Q23020 caenorhabdi
27	28	75.7	6831	5	Q23550	Q23550 caenorhabdi
28	28	75.7	7160	5	Q23551	Q23551 caenorhabdi
29	27	73.0	58	4	Q14489	Q14489 homo sapien
30	27	73.0	89	2	O07069	O07069 burkholderi
31	27	73.0	208	2	O50819	O50819 borrelia bu
32	27	73.0	270	3	O94437	O94437 schizosacch
33	27	73.0	296	3	O43125	O43125 schizosacch
34	27	73.0	310	5	Q20993	Q20993 caenorhabdi
35	27	73.0	332	2	Q9XDU2	Q9xdu2 clostridium
36	27	73.0	335	2	O69440	O69440 leuconostoc
37	27	73.0	342	2	Q9WXG6	Q9wxg6 alcaligenes
38	27	73.0	350	5	Q22946	Q22946 caenorhabdi
39	27	73.0	359	4	O95439	O95439 homo sapien
40	27	73.0	375	4	O15150	O15150 homo sapien
41	27	73.0	389	2	Q48698	Q48698 lactococcus
42	27	73.0	393	6	Q99147	Q99147 oryctolagus
43	27	73.0	400	9	Q9ZX48	Q9zx48 mycobacteri
44	27	73.0	408	5	Q93162	Q93162 caenorhabdi
45	27	73.0	435	4	Q16301	Q16301 homo sapien

ALIGNMENTS

RESULT 1
Q24463
ID Q24463 PRELIMINARY; PRT; 1742 AA.
AC Q24463;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE PROJECTIN (FRAGMENT).
GN BT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE; 93066415.
RA FYRBERG C.C., LABEIT S., BULLARD B., LEONARD K., FYRBERG E.;
RT "Drosophila projectin: relatedness to titin and twitchin and
RT correlation with lethal(4) 102 CDa and bent-dominant mutants.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 249:33-40(1992).
CC -!- FUNCTION: MAY CONFER HIGH RESTING STIFFNESS TO FLIGHT MUSCLES AND
CC MAY ALSO HAVE A ROLE IN ACTIVATING CONTRACTION IN THESE FIBRES IN
CC RESPONSE TO STRETCH.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -!- DEVELOPMENTAL STAGE: FIRST DETECTED AT THE MID-EMBRYO STAGE AND
CC INCREASES PROGRESSIVELY UNTIL THE END OF THE ADULT STAGE.
DR EMBL; X66018; CAA46816.1; -.
DR FLYBASE; FBgn0005666; bt.
DR PFAM; PF00041; fn3; 11.
DR PFAM; PF00047; ig; 1.
KW Muscle protein; Cytoskeleton; Structural protein.
FT NON_TER 1 1
FT NON_TER 1742 1742
SQ SEQUENCE 1742 AA; 192308 MW; 233DED02 CRC32;

Query Match 91.9%; Score 34; DB 5; Length 1742;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFQFRV 7
||:||||

```
Db    292 NEYQFRV 298


RESULT   2
O76281
ID   O76281      PRELIMINARY;      PRT;  6658 AA.
AC   O76281; Q24343;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PROJECTIN (FRAGMENT).
GN   PROJECTIN.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE; 91376068.
RA   AYME-SOUTHGATE A., VIGOREAUX J., BENIAN G., PARDUE M.L.;
RT   "Drosophila has a twitchin/titin-related gene that appears to encode
RT   projectin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:7973-7977(1991).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE; 95146546.
RA   AYME-SOUTHGATE A., SOUTHGATE R., SAIDE J., BENIAN G.M., PARDUE M.L.;
RT   "Both synchronous and asynchronous muscle isoforms of projectin (the
RT   Drosophila bent locus product) contain functional kinase domains.";
RL   J. Cell Biol. 128:393-403(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98300339.
RA   DALEY J., SOUTHGATE R., AYME-SOUTHGATE A.;
RT   "Structure of the Drosophila projectin protein isoforms and
RT   implication for projectin filament assembly.";
RL   J. Mol. Biol. 279:201-210(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   DALEY J.K., SOUTHGATE R.J., AYME-SOUTHGATE A.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF047475; AAC27550.1; -.
DR   HSSP; P00518; 2PHK.
DR   PFAM; PF00041; fn3; 37.
DR   PFAM; PF00047; ig; 8.
DR   PFAM; PF00069; pkinase; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   6658 AA;  743445 MW;  05084AB4 CRC32;

  Query Match             91.9%;  Score 34;  DB 5;  Length 6658;
  Best Local Similarity   85.7%;  Pred. No. 89;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||:||||
Db    2380 NEYQFRV 2386


RESULT   3
Q10466
ID   Q10466      PRELIMINARY;      PRT; 26926 AA.
AC   Q10466;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE; 96026330.
RA   LABEIT S., KOLMER B.;
RT   "Titins: giant proteins in charge of muscle ultrastructure and
RT   elasticity.";
RL   Science 270:293-296(1995).
RN   [2]
RP   SEQUENCE OF 22277-25376 FROM N.A.
RX   MEDLINE; 92258380.
RA   LABEIT S., GAUTEL M., LAKEY A., TRINICK J.;
RT   "Towards a molecular understanding of titin.";
RL   EMBO J. 11:1711-1716(1992).
RN   [3]
RP   SEQUENCE OF 1976-2014 FROM N.A.
RA   LABEIT S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE; 95331314.
RA   GAUTEL M., CASTIGLIONE-MORELLI M.A., PFUHL M., MOTTA A., PASTORE A.;
RT   "A calmodulin-binding sequence in the C-terminus of human cardiac
RT   titin kinase.";
RL   Eur. J. Biochem. 230:752-759(1995).
CC   -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC       ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC       SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
CC       PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
CC       DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
CC       ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
CC       N2-B.
CC   -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC   -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC       KINASES.
CC   -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
CC       IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
CC       DOMAINS.
DR   EMBL; X64698; CAA45939.1; -.
DR   EMBL; X83270; CAA58243.1; -.
DR   EMBL; X64697; CAA45938.1; -.
DR   EMBL; X90568; CAA62188.1; -.
DR   EMBL; X64699; CAA45940.1; -.
DR   HSSP; P56276; 1TLK.
DR   PFAM; PF00041; fn3; 132.
DR   PFAM; PF00047; ig; 59.
DR   PFAM; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00726; LEXASERPTASE.
KW   Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
KW   Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW   Immunoglobulin domain; Phosphorylation.
FT   DOMAIN     1370   1389       4 X 5 AA TANDEM REPEATS OF R-M-S-P-A.
FT   DOMAIN     4429   4614       GLU/LYS/PRO/VAL-RICH.
FT   DOMAIN    24731  25070       CATALYTIC.
FT   DOMAIN    25030  25056       CALMODULIN-BINDING.
FT   MOD_RES    1372   1372       PHOSPHORYLATION (BY PDPK)(POTENTIAL).
FT   MOD_RES    1377   1377       PHOSPHORYLATION (BY PDPK)(POTENTIAL).
FT   MOD_RES    1382   1382       PHOSPHORYLATION (BY PDPK)(POTENTIAL).
FT   MOD_RES    1387   1387       PHOSPHORYLATION (BY PDPK)(POTENTIAL).
FT   MOD_RES   26171  26171       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES   26178  26178       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES   26184  26184       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES   26190  26190       PHOSPHORYLATION (POTENTIAL).
FT   CONFLICT  22277  22277       T -> P (IN REF. 2).
FT   CONFLICT  22449  22449       E -> G (IN REF. 2).
FT   CONFLICT  22454  22454       T -> Q (IN REF. 2).
FT   CONFLICT  23324  23324       S -> L (IN REF. 2).
SQ   SEQUENCE   26926 AA;  2993428 MW;  0214E3A3 CRC32;

  Query Match             91.9%;  Score 34;  DB 4;  Length 26926;
  Best Local Similarity   85.7%;  Pred. No. 3.7e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 NEFQFRV 7
           ||:||||
Db  21623 NEYQFRV 21629


RESULT   4
Q43436
ID   Q43436      PRELIMINARY;      PRT;   294 AA.
AC   Q43436;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   COTTON DNA-BINDING DOMAIN MRNA.
GN   MYB.
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. ACALA SJ-2; TISSUE=OVULE;
RA   WILKINS T.A., LU C.C.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L04497; AAA33067.1; -.
DR   HSSP; P01103; 1POM.
DR   MENDEL; 8407; Goshi;Myb;8407.
DR   PFAM; PF00249; myb_DNA-binding; 2.
KW   DNA-binding.
SQ   SEQUENCE   294 AA;  33202 MW;  1B922EE3 CRC32;


  Query Match             81.1%;  Score 30;  DB 10;  Length 294;
  Best Local Similarity   83.3%;  Pred. No. 28;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           |||||:
Db     169 NEFQFK 174


RESULT   5
Q18810
ID   Q18810      PRELIMINARY;      PRT;   562 AA.
AC   Q18810;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   C53D6.4 PROTEIN.
GN   C53D6.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   GARDNER A.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z70270; CAA94226.1; -.
DR   HSSP; P00274; 1XOA.
DR   PFAM; PF00001; 7tm_1; 1.
SQ   SEQUENCE   562 AA;  64399 MW;  828AB4F4 CRC32;


  Query Match             81.1%;  Score 30;  DB 5;  Length 562;
  Best Local Similarity   83.3%;  Pred. No. 55;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           |||:||
Db     318 NEFEFR 323


RESULT   6
Q22925
ID   Q22925      PRELIMINARY;      PRT;   463 AA.
AC   Q22925;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID C50E3.
GN   C50E3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J , THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L.  WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GEISEL C., BRADSHAW H.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U64848; AAB04880.1; -.
SQ   SEQUENCE   463 AA;  53278 MW;  72E80560 CRC32;


  Query Match             78.4%;  Score 29;  DB 5;  Length 463;
  Best Local Similarity   83.3%;  Pred. No. 75;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFR 6
           |||:||
Db     235 NEFKFR 240


RESULT   7
Q28733
ID   Q28733      PRELIMINARY;      PRT;  6875 AA.
AC   Q28733; Q28736;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
```

```
DE   TITIN (FRAGMENT).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CE12;
RA   LABEIT S.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-6805 FROM N.A.
RC   STRAIN=CE12;
RX   MEDLINE; 92258380.
RA   LABEIT S., GAUTEL M., LAKEY A., TRINICK J.;
RT   "Towards a molecular understanding of titin.";
RL   EMBO J. 11:1711-1716(1992).
RN   [3]
RP   SEQUENCE OF 4305-5320 FROM N.A.
RC   TISSUE=PSOAS MUSCLE;
RX   MEDLINE; 90238553.
RA   LABEIT S., BARLOW D.P., GAUTEL M., GIBSON T., HOLT J., HSIEH C.L.,
RA   FRANCKE U., LEONARD K., WARDALE J., WHITING A., TRINICK J.;
RT   "A regular pattern of two types of 100-residue motif in the sequence
RT   of titin.";
RL   Nature 345:273-276(1990).
DR   EMBL; X64696; CAA45937.1; -.
DR   EMBL; X17329; CAA35207.1; -.
DR   HSSP; P56276; 1TLK.
DR   PFAM; PF00041; fn3; 50.
DR   PFAM; PF00047; ig; 3.
KW   Myosin; Muscle protein.
FT   NON_TER       1      1
FT   NON_TER    6875   6875
SQ   SEQUENCE   6875 AA;  759127 MW;  E5D3B61F CRC32;

  Query Match             78.4%;  Score 29;  DB 6;  Length 6875;
  Best Local Similarity   71.4%;  Pred. No. 1.2e+03;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           :|:|||
Db     396 HEYQFRV 402


RESULT   8
O64232
ID   O64232      PRELIMINARY;      PRT;    66 AA.
AC   O64232;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GP41.1.
GN   41.1.
OS   Mycobacteriophage D29.
OC   Viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FORD M.E., SARKIS G.J., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RL   J. Mol. Biol. 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FORD M.E., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022214; AAC18482.1; -.
SQ   SEQUENCE   66 AA;  7573 MW;  4C0F3242 CRC32;

  Query Match             75.7%;  Score 28;  DB 9;  Length 66;
  Best Local Similarity   83.3%;  Pred. No. 17;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 EFQFRV 7
           |:|:|||
Db      10 EFEFRV 15


RESULT   9
O64191
ID   O64191      PRELIMINARY;      PRT;   120 AA.
AC   O64191;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   HYPOTHETICAL 13.9 KD PROTEIN.
GN   YOTI.
OS   Bacteriophage SPBc2.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LAZAREVIC V., DUESTERHOEFT A., SOLDO B., HILBERT H., MAUEL C.,
RA   KARAMATA D.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF020713; AAC13153.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   120 AA;  13855 MW;  17507F05 CRC32;

  Query Match             75.7%;  Score 28;  DB 9;  Length 120;
  Best Local Similarity   57.1%;  Pred. No. 32;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||| |::
Db      77 NEFNFKI 83


RESULT  10
O30705
ID   O30705      PRELIMINARY;      PRT;   126 AA.
AC   O30705;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YODW (YOTI PROTEIN).
GN   YODW OR YOTI.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WAMBUTT R., WEDLER H., LAPIDUS A., SOROKIN A., EHRLICH D.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 7-126 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
```

```
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 7-126 FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF015775; AAB72080.1; -.
DR   EMBL; Z99114; CAB13878.1; -.
SQ   SEQUENCE   126 AA;  14513 MW;  C52D2056 CRC32;

  Query Match             75.7%;  Score 28;  DB 2;  Length 126;
  Best Local Similarity   57.1%;  Pred. No. 33;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           ||| |::
Db      83 NEFNFKI 89


RESULT  11
O94701
ID   O94701      PRELIMINARY;      PRT;   272 AA.
AC   O94701;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 30.9 KD PROTEIN.
GN   SPBC83.18C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   WOOD V., RAJANDREAM M.A., BARRELL B.G., BOTHE G., POHL T;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035536; CAB36880.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   272 AA;  30855 MW;  F3005AEC CRC32;

  Query Match             75.7%;  Score 28;  DB 3;  Length 272;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQF 5
           |||||
Db     111 NEFQF 115


RESULT  12
O25241
ID   O25241      PRELIMINARY;      PRT;   355 AA.
AC   O25241;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PHOSPHOLIPASE A1 PRECURSOR (DR-PHOSPHOLIPASE A).
GN   HP0499.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000564; AAD07564.1; -.
DR   TIGR; HP0499; -.
KW   Hypothetical protein.
SQ   SEQUENCE   355 AA;  42486 MW;  E73F54E8 CRC32;

  Query Match             75.7%;  Score 28;  DB 2;  Length 355;
  Best Local Similarity   57.1%;  Pred. No. 96;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||:|::
Db     106 NEFKFQI 112


RESULT  13
Q9ZLX5
ID   Q9ZLX5      PRELIMINARY;      PRT;   355 AA.
AC   Q9ZLX5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE PHOSPHOLIPASE A1.
GN   PLDA.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001479; AAD06029.1; -.
SQ   SEQUENCE   355 AA;  42329 MW;  A540BCD9 CRC32;

  Query Match             75.7%;  Score 28;  DB 2;  Length 355;
  Best Local Similarity   57.1%;  Pred. No. 96;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||:|::
Db     106 NEFKFQI 112


RESULT  14
Q23217
ID   Q23217      PRELIMINARY;      PRT;   434 AA.
```

```
AC   Q23217;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   W07G4.2 PROTEIN.
GN   W07G4.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BAYNES C.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z78018; CAB01447.1; -.
SQ   SEQUENCE   434 AA;  50304 MW;  A7BF2973 CRC32;

  Query Match             75.7%;  Score 28;  DB 5;  Length 434;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |:| ||:
Db     331 NDFSFRI 337


RESULT  15
Q25376
ID   Q25376      PRELIMINARY;      PRT;   488 AA.
AC   Q25376;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   VOLTAGE-DEPENDENT POTASSIUM CHANNEL SQKV1A.
OS   Loligo opalescens (California market squid).
OC   Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoida;
OC   Myopsida; Loliginidae; Loligo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MONTEREY BAY, CA; TISSUE=GIANT FIBERLOBE NEURON;
RA   ROSENTHAL J.J.C., VICKERY R., GILLY W.F.;
RL   J. Gen. Physiol. 0:0-0(0).
DR   EMBL; U50543; AAB02884.1; -.
DR   PFAM; PF00520; ion_trans; 1.
DR   PRINTS; PR00169; KCHANNEL.
KW   Ionic channel.
SQ   SEQUENCE   488 AA;  55744 MW;  20E027A5 CRC32;

  Query Match             75.7%;  Score 28;  DB 5;  Length 488;
  Best Local Similarity   85.7%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 NEFQFRV 7
           |||| ||
Db     150 NEFQRRV 156
```

Search completed: June 22, 2000, 19:30:01
Job time: 5143 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 18:02:06 ; Search time 175.2 Seconds
(without alignments)
21.421 Million cell updates/sec

Title: US-09-362-485-28
Perfect score: 15
Sequence: 1 AACGAATTCCGGGTG 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		13.4	89.3	206	1	V04058	Wild-type human ho
2		13.4	89.3	556	1	X33953	Human HCMV inducib
c 3		13.4	89.3	2112	1	V39155	Human glucagon-lik
4		13.4	89.3	3010	1	V52158	Streptococcus pneu
c 5		13.4	89.3	7415	1	X20504	Polynucleotide seq
c 6		13	86.7	207	1	Q89923	Expression vector
c 7		12.4	82.7	48	1	X32592	P. methanolica Aug
c 8		12.4	82.7	48	1	X23879	PCR primer ZC9081.
9		12.4	82.7	153	1	T39087	Partial E. glandul
10		12.4	82.7	1389	1	V21235	Homo sapiens clone
11		12.4	82.7	4062	1	X08434	Inducible nitric o
12		12.4	82.7	4070	1	T98199	Human inducible ni
13		12.4	82.7	4145	1	Q66914	Sequence of the cD
14		12.4	82.7	4145	1	T10115	Nitric oxide synth
15		12.4	82.7	4164	1	Q77700	Nitric-oxide-synth
c 16		12	80.0	20	1	T77593	Wheat microsatelli
17		12	80.0	34	1	V45313	Fv heavy forward p
c 18		12	80.0	35	1	Q69093	Primer for amplify
c 19		12	80.0	40	1	V30727	**Telomerase reverse**
c 20		12	80.0	42	1	V30729	**Telomerase reverse**
c 21		12	80.0	42	1	V30723	Telomerase reverse
c 22		12	80.0	42	1	V30725	Telomerase reverse
c 23		12	80.0	48	1	T71722	Primer 2 for ampli
24		12	80.0	532	1	X22224	Human secreted pro
c 25		12	80.0	736	1	T14225	Bettle patched gen
c 26		12	80.0	736	1	V21586	Tribolium castaneu
c 27		12	80.0	736	1	V64098	Tribolium casteneu
c 28		12	80.0	1617	1	T68031	H. pylori cytoplas
29		12	80.0	1699	1	Q65605	Porcine zona pellu
30		12	80.0	1699	1	V64787	Porcine ZPB cDNA.
31		12	80.0	1919	1	N90817	Rat NADH-cytochrom
c 32		12	80.0	2399	1	V71758	Drosophila mitofus
c 33		12	80.0	3203	1	Q06760	Sequence of beta-g
34		12	80.0	3634	1	V39005	TNF-alpha gene use
35		12	80.0	3634	1	X09014	Tumour necrosis fa
c 36		12	80.0	4946	1	T59531	T-DNA of plasmid p
c 37		12	80.0	5864	1	T39339	Plasmid pTCO113 T-
c 38		12	80.0	6561	1	V33630	Plasmid VLSNO2 DNA
c 39		12	80.0	7699	1	V81410	Plasmid pLAiN used
c 40		12	80.0	7980	1	V81411	Plasmid pLAiH used
41		11.8	78.7	22	1	T67096	Chicken anaemia vi
42		11.8	78.7	24	1	V29308	Mouse liver total
43		11.8	78.7	70	1	V77039	Staphylococcus aur
44		11.8	78.7	190	1	V04057	Wild-type human ho
c 45		11.8	78.7	224	1	T09769	Tomato genomic DNA

ALIGNMENTS

RESULT 1
V04058
ID V04058 standard; DNA; 206 BP.
AC V04058;
DT 05-JUN-1998 (first entry)
DE Wild-type human homologue of RAD9 (HRAD9) genomic DNA fragment.
KW RAD9; yeast checkpoint protein; homologue; radiosensitivity; diagnosis;
KW cancer; radiation; cervical carcinoma; tumour; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 1. .56
FT /*tag= a
FT /number= 1
FT intron 57. .141
FT /*tag= b
FT /number= 1
FT /cons_splice= (5'site: No, 3'site: Yes)
PN WO9746661-A2.
PD 11-DEC-1997.
PF 09-MAY-1997; U08798.
PR 09-MAY-1996; US-644034.
PA (UYCO) UNIV COLUMBIA NEW YORK.
PI Hopkins KM, Lieberman HB;
DR WPI; 98-042172/04.
PT Nucleic acid encoding wild type or mutant human homologue of RAD9
PT yeast checkpoint protein - useful to treat radiosensitivity, and to
PT treat, prevent or diagnose cancer
PS Disclosure; Fig 5C; 56pp; English.
CC This is the genomic DNA fragment of a wild-type human homologue of RAD9
CC (HRAD9), a yeast checkpoint protein. RAD9 is a protein required for cell
CC cycle arrest at the G2 checkpoint in Schizosaccharomyces pombe in
CC response to incomplete replication or damaged DNA. Isolated nucleic acid
CC encoding a novel wild type or mutant human homologue of RAD9 can be used
CC to treat radiosensitive patients, specifically those with ataxia
CC telangiectasia or xeroderma pigmentorum, and prevent cancer in
CC radiosensitive subjects. Wild type RAD9 can be used to treat cancer,
CC especially in patients sensitive to chemotherapy. Oligonucleotides which
CC specifically hybridise to the wild type nucleic acid can be used for
CC ex vivo transformation of tumour cells, which are returned to the
CC patient, who is then treated with chemotherapy and/or radiation,
CC particularly to treat brain, breast, cervical or lung cancer, melanoma
CC **or renal carcinoma. Radiosensitivity can be determined by detecting a**
CC **mutation in the wild type or human mutant homologue of RAD9, useful to**
CC predict the effect of radiotherapy or chemotherapy, and to identify a
CC predisposition to cancer, specifically cervical cancer. The nucleic acid
CC and oligonucleotides specific for the wild type protein can be used to
CC detect the human chromosome region 11q13, while those specific for mutant
CC protein can be used to diagnose cervical carcinoma.
SQ Sequence 206 BP; 34 A; 70 C; 55 G; 47 T;

Query Match 89.3%; Score 13.4; DB 1; Length 206;
Best Local Similarity 93.3%; Pred. No. 60;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 AACGAATTCCGGGTG 15
           || ||||||||||||
```

```
Db      44 AAGGAATTCCGGGTG 58


RESULT   2
X33953
ID   X33953 standard; DNA; 556 BP.
AC   X33953;
DT   30-JUN-1999  (first entry)
DE   Human HCMV inducible gene, SEQ ID NO 28.
KW   HCMV inducible gene; cig; human; human cytomegalovirus; interferon;
KW   anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
KW   drug screening; ds.
OS   Homo sapiens.
PN   WO9913075-A2.
PD   18-MAR-1999.
PF   08-SEP-1998; U18638.
PR   22-SEP-1997; US-059725.
PR   08-SEP-1997; US-058180.
PA   (UYPR-) UNIV PRINCETON.
PI   Cong J, Schenk T, Zhu H;
DR   WPI; 99-243729/20.
DR   P-PSDB; Y05378.
PT   New isolated human genes
PS   Claim 2; Page 160-161; 184pp; English.
CC   This sequence represents a human gene of the invention, that is induced
CC   to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC   genes (cig or cigs). The invention also relates to genes that are
CC   repressed in the presence of HCMV infection, designated HCMV-repressible
CC   genes (crg or crgs). The products can be used to obtain agents which can
CC   be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC   also be used for the development of drugs that would allow for higher
CC   dosage IFN treatments without the concomitant toxicity normally
CC   associated with administering high levels of IFN. The products can also
CC   be used for detection, diagnosis and drug screening.
SQ   Sequence   556 BP;    112 A;    160 C;    198 G;    86 T;

  Query Match             89.3%;  Score 13.4;  DB 1;  Length 556;
  Best Local Similarity   93.3%;  Pred. No. 59;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||| ||||
Db      37 AACGAATTCCAGGTG 51


RESULT   3
V39155/c
ID   V39155 standard; cDNA; 2112 BP.
AC   V39155;
DT   23-SEP-1998  (first entry)
DE   Human glucagon-like peptide-2 (GLP-2) receptor encoding cDNA.
KW   Glucagon-like peptide-2 receptor; GLP-2 receptor; human; rat;
KW   receptor ligand screening; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             122. .1783
FT                   /*tag= a
FT                   /product= "human GLP-2 receptor"
FT   sig_peptide     122. .319
FT                   /*tag= b
FT   mat_peptide     320. .1780
FT                   /*tag= c
PN   WO9825955-A2.
PD   18-JUN-1998.
PF   15-DEC-1997; CA0969.
PR   24-APR-1997; US-845546.
PR   13-DEC-1996; US-767224.
PR   24-JAN-1997; US-787721.
PA   (ALLX ) ALLELIX BIOPHARMACEUTICALS INC.
PI   Fan E, Gupta AK, McCallum K, Munroe DG, Vyas TB;
DR   WPI; 98-348449/30.
DR   P-PSDB; W68065.
PT   New, e.g. rat glucagon-like peptide-2 receptor - useful for, e.g.
PT   isolating homologues and GLP-2 receptor ligand screening
PS   Claim 3; Fig 6A-B; 54pp; English.
CC   This cDNA encodes a human glucagon-like peptide-2 (GLP-2) receptor. The
CC   specification provides the sequences of human and rat GLP-2 receptors.
CC   Genetically engineered host cells containing the GLP-2 receptor encoding
CC   nucleic acid sequences operably linked to expression elements can be
CC   used to produce the recombinant proteins. These cells can be used in a
CC   method for identifying GLP-2 receptor ligands which comprises incubating
CC   a candidate ligand with the cell, and determining whether binding has
CC   occurred between the GLP-2 receptor and the candidate ligand. The GLP-2
CC   receptor polynucleotides can be used to isolate GLP-2 receptor-encoding
CC   homologues. They can also be used for constructing cell lines for GLP-2
CC   receptor ligand screening.
SQ   Sequence   2112 BP;    470 A;    567 C;    545 G;    530 T;

  Query Match             89.3%;  Score 13.4;  DB 1;  Length 2112;
  Best Local Similarity   93.3%;  Pred. No. 58;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           || ||||||||||||
Db    2088 AAAGAATTCCGGGTG 2074


RESULT   4
V52158
ID   V52158 standard; DNA; 3010 BP.
AC   V52158;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:25.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI; 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 294-296; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence   3010 BP;    907 A;    535 C;    716 G;    852 T;
```

```
  Query Match             89.3%;  Score 13.4;  DB 1;  Length 3010;
  Best Local Similarity   93.3%;  Pred. No. 58;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           ||||||||||||  ||
Db     652 AACGAATTCCGGCTG 666


RESULT   5
X20504/c
ID   X20504 standard; DNA; 7415 BP.
AC   X20504;
DT   05-MAY-1999  (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum.
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI; 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1; Page 202-206; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
SQ   Sequence   7415 BP;    1536 A;    1564 C;    2199 G;    2109 T;


  Query Match             89.3%;  Score 13.4;  DB 1;  Length 7415;
  Best Local Similarity   93.3%;  Pred. No. 58;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||  ||||||||||||
Db    3879 AACAAATTCCGGGTG 3865


RESULT   6
Q89923/c
ID   Q89923 standard; DNA; 207 BP.
AC   Q89923;
DT   04-DEC-1995  (first entry)
DE   Expression vector TrpLEPTH fragment.
KW   Analogue; truncated human parathyroid hormone; PTH; hPTH; substitution;
KW   osteoporosis; hypercalcaemia; hyperparathroidism;
KW   metabolic bone disease; human; veterinary medicine;
KW   iontophoretic transdermal transport; recombinant E.coli; ss.
OS   Synthetic.
PN   WO9511988-A.
PD   04-MAY-1995.
PF   25-OCT-1994; U12205.
PR   25-OCT-1993; US-142551.
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
PI   Oldenburg KR,  Selick HE;
DR   WPI; 95-178880/23.
DR   P-PSDB; R74528.
PT   New active analogues of parathyroid hormone - with increased
PT   activity, stability in serum etc., esp. for treating
PT   osteoporosis, also related DNA and vectors
PS   Disclosure; Fig 3; 109pp; English.
CC   This sequence represents a fragment of the expression vector, TrpLEPTH.
CC   This vector encodes a fusion protein comprised of a leader peptide
CC   sequence, residues 1-34 of human parathyroid hormone (PTH) and 4-6 His
CC   residues that serve as a tag for purification of the protein on a nickel
CC   column.  This sequence was used in the production of analogues of the
CC   truncated form of PTH.  These analogues have increased activity and
CC   longer serum half life than native PTH due to eg. substitution of Met
CC   residues with Leu residues and replacing the carboxy Phe with Tyr.  The
CC   carboxy terminal may also be modified by the addition of a homoserine
CC   residue or analogue, or by the addition of residues 35-84 of wild type
CC   PTH (see R74410).  These PTH analogues may be used in the treatment of
CC   osteoporosis or hypercalcaemia, hyperparathroidism or other metabolic
CC   bone diseases in human or veterinary medicine.  These peptides may also
CC   have increased iontophoretic transdermal transport compared to wild type
CC   PTH and can be produced in high yield in recombinant E.coli.
SQ   Sequence   207 BP;    59 A;    62 C;    40 G;    46 T;


  Query Match             86.7%;  Score 13;  DB 1;  Length 207;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCGGGT 14
           |||||||||||||
Db      53 ACGAATTCCGGGT 41


RESULT   7
X32592/c
ID   X32592 standard; DNA; 48 BP.
AC   X32592;
DT   24-JUN-1999  (first entry)
DE   P. methanolica Aug1 gene specific primer ZC9081.
KW   Pichia methanolica; vacuolar protease; genetic defect; proteinase;
KW   industrial; pharmaceutical; enzyme; enzyme inhibitor; growth factor;
KW   cytokine; hormone; insulin; leptin; glucagon; proteolysis; Aug1;
KW   PCR primer; ss.
OS   Synthetic.
OS   Pichia methanolica.
PN   WO9914347-A1.
PD   25-MAR-1999.
PF   11-SEP-1998; U19449.
PR   15-SEP-1997; US-929748.
PA   (ZYMO ) ZYMOGENETICS INC.
PI   Raymond CK, Vanaja E;
DR   WPI; 99-244037/20.
PT   Pichia methanolica defective in vacuolar protease
PS   Example 5; Page 51; 54pp; English.
CC   The invention relates to Pichia methanolica cell which are functionally
CC   defective in a vacuolar protease. The functional defeciency is due to a
CC   genetic defect in the parent gene, wherein said parent gene encodes
CC   proteinase A or proteinase B. The new cells are hosts for production of
CC   heterologous proteins of industrial or pharmaceutical value, e.g. a wide
CC   range of enzymes, enzyme inhibitors, growth factors, cytokines and
CC   hormones (such as insulin, leptin and glucagon). The protease-defective
CC   cells should show reduced proteolysis of recombinant proteins. Sequences
CC   X32582-595 represent PCR primers used during the course of the invention.
SQ   Sequence   48 BP;    19 A;    8 C;    8 G;    13 T;


  Query Match             82.7%;  Score 12.4;  DB 1;  Length 48;
  Best Local Similarity   92.9%;  Pred. No. 2.2e+02;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGT 14
           ||||||||||||| |
Db      21 AACGAATTCCGGAT 8


RESULT   8
X23879/c
ID   X23879 standard; DNA; 48 BP.
AC   X23879;
DT   25-JUN-1999  (first entry)
DE   PCR primer ZC9081.
KW   Chromosomal mutagenesis; chromosomal loci; genotype; enzyme production;
```

KW lipase; cellulase; protease; enzyme inhibitor; growth factor; cytokine;
KW platelet derived growth factor; fibroblast growth factor; hormone;
KW epidermal growth factor; erythropoietin; thrombopoietin; insulin; leptin;
KW glucagon; PCR primer; ss.
OS Synthetic.
OS Pichia methanolica.
PN WO9914320-A1.
PD 25-MAR-1999.
PF 11-SEP-1998; U19448.
PR 30-DEC-1997; US-001141.
PR 15-SEP-1997; US-929748.
PA (ZYMO) ZYMOGENETICS INC.
PI Raymond CK;
DR WPI; 99-229528/19.
PT Altering a selected chromosomal locus to produce strains with
PT desired genotypes
PS Example 5; Page 68; 72pp; English.
CC This invention describes a novel method for introducing mutations into
CC chromosomal loci of Pichia methanolica to produce strains having
CC desired genotypes. P. methanolica strains having altered target loci
CC are useful as hosts for the expression of heterologous genes. Proteins
CC that can be produced in P. methanolica included proteins of industrial
CC and pharmaceutical interest, e.g. enzymes (lipases, cellulases,
CC proteases), enzyme inhibitors, growth factors such as a platelet derived
CC growth factor, fibroblast growth factor and epidermal growth factor,
CC cytokines such as erythropoietin and thrombopoietin, and hormones such as
CC insulin, leptin and glucagon. Directed mutagenesis allows the
CC introduction of mutations into predetermined genomic loci, permitting the
CC selective alteration of gene activity. However, some mutation methods are
CC unsuitable for Pichia methanolica cells, e.g. the pop-in/pop-out method
CC which requires selection against 5-fluoro orotic acid to which
CC P. methanolica cells are resistant. The methods can be used to easily
CC and readily mutate P. methanolica cells.
SQ Sequence 48 BP; 19 A; 8 C; 8 G; 13 T;

Query Match 82.7%; Score 12.4; DB 1; Length 48;
Best Local Similarity 92.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACGAATTCCGGGT 14
||||||||||||| |
Db 21 AACGAATTCCGGAT 8

RESULT 9
T39087
ID T39087 standard; cDNA; 153 BP.
AC T39087;
DT 21-MAY-1997 (first entry)
DE Partial E. glandulosa endoglucanase coding sequence.
KW Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;
KW plant cellulase; catalytic region; textile; backstaining; bio-polishing;
KW stone-washing; cellulosic fabric; colour clarification; defibration;
KW cell wall degradation; paper pulp; debarking; fibre modification;
KW enzymatic de-inking; drainage improvement; ss.
OS Exidia glandulosa.
PN WO9629397-A1.
PD 26-SEP-1996.
PF 18-MAR-1996; DK0105.
PR 17-MAR-1995; DK-000272.
PR 08-AUG-1995; DK-000887.
PR 08-AUG-1995; DK-000885.
PR 08-AUG-1995; DK-000888.
PR 08-AUG-1995; DK-000886.
PR 12-FEB-1996; DK-000137.
PA (NOVO) NOVO-NORDISK AS.
PI Andersen LN, Ihara M, Kauppinen MS, Lange L, Lassen SF;
PI Nielsen RI, Schuelein M, Takagi S;
DR WPI; 96-443173/44.
DR P-PSDB; W04949.
PT New endo:glucanase enzyme preparations - contg. conserved catalytic
PT regions, useful for treating fabrics, textiles, plant material or
PT paper pulp
PS Example 5; Page 209-210; 316pp; English.
CC T39076-T39100, and T41491 represent partial endoglucanase coding
CC sequences. The encoded peptides are examples of fragments of the enzymes
CC of the invention (see W04925 for an example of a full length
CC endoglucanase of the invention). The enzymes possess cellulytic
CC (particularly endoglucanase) activity. Cellulytic enzymes are involved
CC in the hydrolysis of cellulose, and are synthesised by a large number of
CC microorganisms and plants. The enzymes of the invention containing the
CC conserved catalytic regions (such as W04913) exhibit improved
CC performance, e.g. 50 times higher performance, compared to multiple
CC domain enzymes. The enzymes can be used for the treatment of fabrics or
CC textiles, preferably for preventing backstaining, for bio-polishing or
CC for stone-washing cellulosic fabric. They can also be used to provide
CC colour clarification for laundry. The enzymes can also be used for the
CC degradation or modification of plant material, such as cell walls. They
CC can also be used in the treatment of paper pulp preferably for debarking,
CC defibration, fibre modification, enzymatic de-inking or drainage
CC improvement.
SQ Sequence 153 BP; 40 A; 38 C; 42 G; 33 T;

Query Match 82.7%; Score 12.4; DB 1; Length 153;
Best Local Similarity 92.9%; Pred. No. 2.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ACGAATTCCGGGTG 15
||||| |||||||||
Db 37 ACGAAATCCGGGTG 50

RESULT 10
V21235
ID V21235 standard; cDNA; 1389 BP.
AC V21235;
DT 03-AUG-1998 (first entry)
DE Homo sapiens clone D329_1.
KW secreted protein; human; PBMC; peripheral blood mononuclear cells; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1042..1389
FT /*tag= a
FT /product= D329_1 protein
PN WO9807859-A2.
PD 26-FEB-1998.
PF 22-AUG-1997; U14874.
PR 23-AUG-1996; US-702344.
PA (GEMY) Genetics Inst Inc.
PI Merberg F, McCoy JM, Lavallie ER, Racie LA, Treacy M, Spaulding V,
PI Jacobs K;
DR WPI; 98-169163/15.
DR P-PSDB; W52989.
PT New nucleic acid encoding secreted proteins from human cells -
PT useful e.g. as immuno-modulators, antitumour agents, promoters of
PT tissue growth, haemostatic and thrombolytic agents etc.
PS Claim 28; Pages 58-59; 79pp; English.
CC The sequence is that of the clone D329_1 which was
CC isolated from a human adult PBMC cDNA library using methods
CC selective for cDNAs that encode secreted proteins.
SQ Sequence 1389 BP; 272 A; 402 C; 355 G; 360 T;

Query Match 82.7%; Score 12.4; DB 1; Length 1389;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ACGAATTCCGGGTG 15
||| ||||||||||
Db 431 ACGGATTCCGGGTG 444

RESULT 11
X08434

```
ID   X08434 standard; DNA; 4062 BP.
AC   X08434;
DT   28-JUN-1999  (first entry)
DE   Inducible nitric oxide synthase gene.
KW   Manganese containing superoxide dismutase; MnSOD; IDDM;
KW   diabetes mellitus; treatment; therapy; nitric oxide; NO; beta cell;
KW   fatty acids; lipotoxic; cytotoxic; cytokine; osteoporosis;
KW   inflammatory disease; autoimmune disease; neurodegenerative disease;
KW   ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             106. .3567
FT                   /*tag= a
FT                   /product= "Nitric oxide synthase"
PN   WO9906059-A2.
PD   11-FEB-1999.
PF   30-JUL-1998; U15781.
PR   03-MAR-1998; US-055092.
PR   30-JUL-1997; US-055092.
PA   (BETA-) BETAGENE INC.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Clark SA, Hohmeier H, Koyama K, Lee Y, Newgard CB,
PI   Ohneda M, Shimabukurom, Thigpen A, Unger RH;
DR   WPI; 99-153448/13.
DR   P-PSDB; W96322.
PT   Protection of mammalian cells against immunotoxicity or lipotoxicity
PT   - used for treating, e.g. diabetes, obesity, wasting syndromes,
PT   osteoporosis, inflammatory diseases, autoimmune diseases or
PT   neurodegenerative diseases
PS   Disclosure; Page 244-247; 253pp; English.
CC   Inhibition of cytokine mediated immunotoxicity of cells can be
CC   achieved by blocking free radical production or the accumulation of
CC   free radicals in that cell. Treatment of insulin dependent diabetes
CC   mellitus (IDDM) can be achieved by by blocking nitric oxide (NO)
CC   production in a pancreatic beta cell and by providing a composition
CC   comprising an agent that reduces levels of fatty acids in the cells
CC   and protects beta-cells of the subject against lipid-mediated cell
CC   death.  Cells can also be protected against nitric oxide mediated
CC   cytotoxicity by introducing into the cell an antioxidising agent.
CC   The methods can be used for protecting cells against immunotoxicity
CC   mediated by, e. g. 1L-1 beta, 1L-1 alpha, gamma IFN, TNF alpha, TNF
CC   beta, IL-8, IL-2, IL-6, IL-2, IL-3, IL-5, IL-7, IL-9, IL-14, IL-17,
CC   granulocyte-macrophage colony stimulating factor or monocyte
CC   chemoattractant protein-1. The methods can be used for the treatment
CC   of e.g.  insulin-dependent diabetes mellitus (IDDM), NIDDM, obesity,
CC   wasting syndromes, short stature, osteoporosis, inflammatory
CC   diseases, autoimmune diseases, or neurodegenerative diseases.
SQ   Sequence   4062 BP;    966 A;    1178 C;    1099 G;    819 T;

  Query Match             82.7%;  Score 12.4;  DB 1;  Length 4062;
  Best Local Similarity   92.9%;  Pred. No. 2.1e+02;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCGGGTG 15
           |||| ||||||||||
Db     869 ACGACTTCCGGGTG 882


RESULT  12
T98199
ID   T98199 standard; cDNA; 4070 BP.
AC   T98199;
DT   11-MAY-1998  (first entry)
DE   Human inducible nitric oxide synthetase cDNA.
KW   Nitric oxide synthetase; NOS; iNOS; HPiNOS; human;
KW   erectile dysfunction; impotence; gene therapy; corpora cavernosa;
KW   relaxant; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             197. .3658
FT                   /*tag= a
PN   WO9742965-A1.
PD   20-NOV-1997.
PF   09-MAY-1997; U07643.
PR   10-MAY-1996; US-017373.
PA   (GONZ/) GONZALEZ-CADAVID N F.
PA   (RAJF/) RAJFER J.
PI   Gonzalez-Cadavid NF, Rajfer J;
DR   WPI; 98-008577/01.
DR   P-PSDB; W36113.
PT   Treatment of erectile dysfunction - by introducing an agent into
PT   penile tissue, particularly for inducing cavernosal smooth muscle
PT   relaxation or increasing NOS levels
PS   Claim 14; Page 35-38; 53pp; English.
CC   This cDNA sequence includes a coding region for human penis
CC   inducible nitric oxide synthetase (HPiNOS) (see W36113).  It was
CC   isolated by reverse transcription of mRNA from human penile smooth
CC   muscle cells with iNOS antisense primers, combined with PCR
CC   amplification of the resulting cDNA fragments.  The invention is
CC   directed to a method of treating erectile dysfunction in a
CC   patient by providing an agent capable of treating erectile
CC   dysfunction, and introducing an effective amount of the agent into
CC   the penile tissue of the patient.  Preferably, the agent induces
CC   cavernosal smooth muscle relaxation, and/or produces an increase
CC   in the level of NOS in tissue.  Preferably, the NOS is iNOS, and
CC   the agent is introduced into the corpora cavernosa of the penis.
CC   The agent is preferably an NOS inducer, an NOS protein such as
CC   HPiNOS, a cDNA encoding an NOS such as HPiNOS, or cDNA-transformed
CC   penile cells, especially corpora cavernosa cells.
SQ   Sequence   4070 BP;    947 A;    1197 C;    1112 G;    814 T;

  Query Match             82.7%;  Score 12.4;  DB 1;  Length 4070;
  Best Local Similarity   92.9%;  Pred. No. 2.1e+02;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 ACGAATTCCGGGTG 15
           |||| ||||||||||
Db     960 ACGACTTCCGGGTG 973


RESULT  13
Q66914
ID   Q66914 standard; cDNA; 4145 BP.
AC   Q66914;
DT   28-DEC-1994  (first entry)
DE   Sequence of the cDNA clone for human hepatocyte inducible nitric
DE   oxide synthase.
KW   Nitric oxide synthase; hepatocyte; hypotensive shock; therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             207. .3668
FT                   /*tag= a
PN   WO9412645-A.
PD   09-JUN-1994.
PF   23-NOV-1993; U11401.
PR   25-NOV-1992; US-981344.
PA   (UYPI-) UNIV PITTSBURGH.
PI   Billiar TR,  Geller DA,  Nussler AK,  Simmons RL;
DR   WPI; 94-200273/24.
DR   P-PSDB; R55764.
PT   cDNA clone encoding human inducible nitric oxide synthase - used
PT   to prevent the hypotensive shock seen with sepsis.
PS   Claim 23; Fig 1; 53pp; English.
CC   Q66914 is from human hepatocyte inducible nitric oxide synthase cDNA
CC   clone pHINOS from lambda Zap II cDNA library. The original source
CC   was induced human hepatocyte RNA. HINOS cDNA plasmid is pref.
CC   transformed in E. coli SOLR (ATCC 69126). The inventors claim a
CC   clone with the cDNA sequence in Q66914 and a cDNA clone which
CC   encodes R55764. The cloning and expression of a human tissue nitric
CC   oxide synthase cDNA provides a source of the enzyme for therapeutic
CC   purposes, for example to prevent the hypotensive shock seen with
CC   sepsis.
SQ   Sequence   4145 BP;    968 A;    1203 C;    1126 G;    848 T;
```

Query Match 82.7%; Score 12.4; DB 1; Length 4145;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 ACGAATTCCGGGTG 15
          |||| |||||||||
Db    970 ACGACTTCCGGGTG 983
```

RESULT 14
T10115
ID T10115 standard; cDNA; 4145 BP.
AC T10115;
DT 13-MAY-1996 (first entry)
DE Nitric oxide synthase cDNA clone pHINOS.
KW Inducible nitric oxide synthase; iNOS; hepatocyte; gene therapy;
KW vascular occlusive disease; cancer; infection; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 207. .3668
FT /*tag= a
PN WO9600006-A1.
PD 04-JAN-1996.
PF 20-JUN-1995; U07849.
PR 24-JUN-1994; US-265046.
PA (UYPI-) UNIV PITTSBURGH.
PI Billiar TR, Geller DA, Nussler AK, Simmons RL, Tzeng E;
DR WPI; 96-068641/07.
DR P-PSDB; R88464.
PT Inducible nitric oxide synthase gene - useful in gene therapy to
PT treat, e.g. vascular occlusive disease and cancer
PS Claim 72; Page 53-58; 91pp; English.
CC A cDNA clone (T10115), designated pHINOS, codes for the human
CC hepatocyte inducible nitric oxide synthase (iNOS = R88464). It was
CC obtd. by isolating mRNA from hepatocytes induced in vitro for iNOS
CC biosynthesis, preparing a cDNA library in a phage lambda Zap II
CC vector, and screening with a cross-species iNOS probe. The cDNA can
CC be used to prepare iNOS for therapeutic use. Alternatively, it is
CC used in gene therapy strategies for treatment of vascular occlusive
CC disease associated with atherosclerosis, vascular bypass and diabetes
CC mellitus, tumor cell growth, and microbial infections.
SQ Sequence 4145 BP; 968 A; 1205 C; 1124 G; 848 T;

Query Match 82.7%; Score 12.4; DB 1; Length 4145;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 ACGAATTCCGGGTG 15
          |||| |||||||||
Db    970 ACGACTTCCGGGTG 983
```

RESULT 15
Q77700
ID Q77700 standard; cDNA; 4164 BP.
AC Q77700;
DT 09-MAY-1995 (first entry)
DE Nitric-oxide-synthase pBSHSINOS clone.
KW Nitric-oxide-synthase; NO-synthase; NOS; chondrocyte;
KW interleukin-1-beta; pBSHSINOS; arthritis; hypertension;
KW septic shock; inflammation; ischemia; dementia; obesity; tumor;
KW agonist; antagonist; vector; CHO; Chinese hamster ovary;
KW cell culture; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 226. .3687
FT /*tag= a
PN WO9423038-A.
PD 13-OCT-1994.
PF 25-MAR-1994; G00621.
PR 26-MAR-1993; GB-006386.
PA (WELL) WELLCOME FOUND LTD.
PI Charles IG, Moncada SE, Palmer RMJ, Moncada S;
DR WPI; 94-333198/41.
DR P-PSDB; R63206.
PT New human inducible nitric oxide synthase - useful for
PT identifying enzyme inhibitors and stimulators, and for diagnosis
PT and treatment of e.g. viral infections or tumours
PS Disclosure; Page 25-31; 42pp; English.
CC Human chondrocytes were incubated with interleukin-1-beta to induce
CC nitric-oxide-synthase. cDNA was generated and used to construct a
CC library in lambda ZAPII. This was screened with a 650 bp fragment of
CC mouse inducible NO-synthase cDNA to identify the full-length clone
CC pBSHSINOS. The insert was transferred to pSVL to give a vector
CC capable of expressing NO-synthase in CHO cells under control of a
CC heterologous constitutive promoter.
SQ Sequence 4164 BP; 974 A; 1210 C; 1127 G; 853 T;

Query Match 82.7%; Score 12.4; DB 1; Length 4164;
Best Local Similarity 92.9%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 ACGAATTCCGGGTG 15
          |||| |||||||||
Db    989 ACGACTTCCGGGTG 1002
```

Search completed: June 22, 2000, 18:02:09
Job time: 10141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2000, 17:43:49 ; Search time 2113.77 Seconds
(without alignments)
28.763 Million cell updates/sec

Title: US-09-362-485-28
Perfect score: 15
Sequence: 1 AACGAATTCCGGGTG 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45 *
80: gb_est46:*
81: gb_est47:*
82: gb_gss1:*
83: gb_gss2:*
84: gb_gss3:*
85: gb_gss4:*
86: em_gss1:*
87: em_gss2:*
88: em_gss3:*
89: em_gss4:*
90: gb_gss5:*
91: gb_gss6:*
92: gb_gss7:*
93: gb_gss8:*
94: gb_gss9:*
95: em_gss5:*
96: em_gss6:*
97: em_gss7:*
98: em_gss8:*
99: em_gss9:*
100: em_gss10:*
101: em_gss11:*
102: gb_gss10:*
103: gb_gss11:*
104: em_gss12:*
105: gb_gss12:*
106: gb_gss13:*
107: gb_gss14:*
108: gb_gss15:*
109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	% Query

	No.	Score	Match	Length	DB	ID	Description
	1	14	93.3	242	30	AA257811	AA257811 MBL2SJ3G1
	2	14	93.3	243	35	AA542789	AA542789 MBL2SK3F0
	3	14	93.3	524	69	AV386258	AV386258 AV386258
c	4	14	93.3	543	39	AA898689	AA898689 NCC3A9T7
c	5	14	93.3	871	108	AQ574334	AQ574334 nbxb0085G
	6	13.4	89.3	160	59	AV081419	AV081419 AV081419
	7	13.4	89.3	182	25	W05084	W05084 za79b12.r1
	8	13.4	89.3	209	20	T28507	T28507 EST47206 Hu
c	9	13.4	89.3	210	80	AW392780	AW392780 CM3-ST027
c	10	13.4	89.3	234	64	AL036722	AL036722 DKFZp564G
c	11	13.4	89.3	240	35	C45133	C45133 C45133 Yuji
c	12	13.4	89.3	248	38	AA767764	AA767764 ob37d05.s
c	13	13.4	89.3	256	23	H41052	H41052 yp52c07.s1
c	14	13.4	89.3	256	25	N90615	N90615 zb42h08.s1
	15	13.4	89.3	262	26	W24444	W24444 zb60b04.r1
c	16	13.4	89.3	277	25	N90597	N90597 zb42f02.s1
c	17	13.4	89.3	280	63	AI970480	AI970480 wq93d05.x
c	18	13.4	89.3	292	49	AI627641	AI627641 ty81b02.x
c	19	13.4	89.3	298	49	AI627652	AI627652 ty81c04.x
c	20	13.4	89.3	298	51	AI759983	AI759983 wh36f10.x
c	21	13.4	89.3	300	50	AI681913	AI681913 tx50e11.x
c	22	13.4	89.3	301	64	AW044272	AW044272 wy71h01.x
c	23	13.4	89.3	307	46	AI459664	AI459664 ar84a12.x
c	24	13.4	89.3	316	21	F10945	F10945 HSC3MB072 n
	25	13.4	89.3	331	29	AA165585	AA165585 zo81h01.r
c	26	13.4	89.3	334	23	H44202	H44202 yo73g11.s1
c	27	13.4	89.3	339	79	AW236081	AW236081 xn24g11.x
c	28	13.4	89.3	352	35	AA558280	AA558280 nl54a02.s
c	29	13.4	89.3	360	35	C48699	C48699 C48699 Yuji
c	30	13.4	89.3	367	25	N69276	N69276 za25b04.s1
c	31	13.4	89.3	370	23	H50689	H50689 yp10b03.s1
	32	13.4	89.3	375	28	AA084812	AA084812 zn02c09.r
c	33	13.4	89.3	385	26	W81396	W81396 zd86g07.s1
	34	13.4	89.3	393	28	AA112437	AA112437 zm27f06.r
c	35	13.4	89.3	397	24	H92861	H92861 yt91b12.s1
	36	13.4	89.3	402	22	R82716	R82716 yj20f08.r1
c	37	13.4	89.3	404	61	AI824882	AI824882 wb03c03.x
c	38	13.4	89.3	405	22	R60905	R60905 yh05b08.s1
c	39	13.4	89.3	407	38	AA768399	AA768399 oa63b09.s
c	40	13.4	89.3	409	35	AA574090	AA574090 nl77f01.s
c	41	13.4	89.3	410	22	R37795	R37795 yf64e06.s1
c	42	13.4	89.3	414	22	R38192	R38192 yc93f07.s1
	43	13.4	89.3	418	28	AA076221	AA076221 zm18e07.r
	44	13.4	89.3	418	29	AA191138	AA191138 zp86c12.r
	45	13.4	89.3	421	91	AQ850908	AQ850908 LMAJFV1_1

ALIGNMENTS

RESULT 1
AA257811
LOCUS AA257811 242 bp mRNA EST 22-SEP-1997
DEFINITION MBL2SJ3G11T3 Brugia malayi second stage larvae JHU96SL-BmL2 Brugia malayi cDNA clone L2SJ3G11 5', mRNA sequence.
ACCESSION AA257811
VERSION AA257811.1 GI:1893327
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 242)
AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones,S.J.
TITLE Genes expressed in adult female Brugia malayi
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405261.
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at http://www.sanger.ac.uk/brugia/L2S/MBL2SJ3G11T3.html This is the full sequence of the cDNA clone. The polyA tail has been clipped and is excluded from this sequence
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .242
/organism="Brugia malayi"
/strain="TRS Labs"
/db_xref="taxon:6279"
/clone="L2SJ3G11"
/clone_lib="Brugia malayi second stage larvae JHU96SL-BmL2"
/sex="mixed"
/dev_stage="mosquito derived, second stage larvae (L2)"
/lab_host="E. coli XL1-Blue MRF-"
/note="Vector: lambdaZapII; Site_1: Eco R I (5' end); Site_2: Xho I (3' end); Brugia malayi is a lymphatic filarial nematode parasite of humans. Full length cDNA was prepared by long-range RT-PCR from mRNA from L2 larvae of the human filarial nematode parasite Brugia malayi using nematode spliced leader (SL, 5'end) and oligo-d(T) (3' end) primers. The library had an unamplified titre of ~1 x 10E6 per ml and ~95% of clones have inserts (mean length ~900 bp). The library is available from The Filarial Genome Project Resource Center: contact Dr. S.A. Williams, Clark Science Center, Smith College, Northampton, MA 01063 USA phone +1 413 585 3826 fax +1 413 585 3786 email genome@smith.edu"
BASE COUNT 71 a 43 c 58 g 70 t
ORIGIN

Query Match 93.3%; Score 14; DB 30; Length 242;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AACGAATTCCGGGT 14
           ||||||||||||||
Db     207 AACGAATTCCGGGT 220
```

RESULT 2
AA542789
LOCUS AA542789 243 bp mRNA EST 22-SEP-1997
DEFINITION MBL2SK3F05T3 Brugia malayi second stage larvae JHU96SL-BmL2 Brugia malayi cDNA clone L2SK3F05 5', mRNA sequence.
ACCESSION AA542789
VERSION AA542789.1 GI:2289701
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 243)
AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones,S.J.
TITLE Genes expressed in adult female Brugia malayi
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1401020.
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

```
          The ABI trace of this sequence can be viewed at
          http://www.sanger.ac.uk/brugia/L2S/MBL2SK3F05T3.html This is the
          full sequence of the cDNA clone. The polyA tail has been clipped
          and is excluded from this sequence
          Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Brugia malayi"
                     /strain="TRS Labs"
                     /db_xref="taxon:6279"
                     /clone="L2SK3F05"
                     /clone_lib="Brugia malayi second stage larvae
                     JHU96SL-BmL2"
                     /sex="mixed"
                     /dev_stage="mosquito derived, second stage larvae (L2)"
                     /lab_host="E. coli XL1-Blue MRF-"
                     /note="Vector: lambdaZapII; Site_1: Eco R I (5' end);
                     Site_2: Xho I (3' end); Brugia malayi is a lymphatic
                     filarial nematode parasite of humans. Full length cDNA was
                     prepared by long-range RT-PCR from mRNA from L2 larvae of
                     the human filarial nematode parasite Brugia malayi using
                     nematode spliced leader (SL, 5'end) and oligo-d(T) (3'
                     end) primers. The library had an unamplified titre of ~1 x
                     10E6 per ml and ~95% of clones have inserts (mean length
                     ~900 bp). The library is available from The Filarial
                     Genome Project Resource Center: contact Dr. S.A. Williams,
                     Clark Science Center, Smith College, Northampton, MA 01063
                     USA phone +1 413 585 3826 fax  +1 413 585 3786 email
                     genome@smith.edu"
BASE COUNT       71 a     43 c     58 g     71 t
ORIGIN


  Query Match             93.3%;  Score 14;  DB 35;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGT 14
           ||||||||||||||
Db     207 AACGAATTCCGGGT 220


RESULT   3
AV386258
LOCUS       AV386258      524 bp    mRNA            EST       27-OCT-1999
DEFINITION  AV386258 Halocynthia roretzi Fertilized egg Halocynthia roretzi
            cDNA clone 008L09_3 3', mRNA sequence.
ACCESSION   AV386258
VERSION     AV386258.1  GI:6131315
KEYWORDS    EST.
SOURCE      Halocynthia roretzi.
  ORGANISM  Halocynthia roretzi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Makabe,K.W.
  TITLE     Halocynthia roretzi EST
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5866246.
            Contact: Kazuhiro W. Makabe
            Department of Zoology, Graduate School of Science
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Halocynthia roretzi"
                     /db_xref="taxon:7729"
                     /clone="008L09_3"
                     /clone_lib="Halocynthia roretzi Fertilized egg"
                     /dev_stage="Fertilized egg"
BASE COUNT      187 a     88 c     96 g    153 t
ORIGIN


  Query Match             93.3%;  Score 14;  DB 69;  Length 524;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGT 14
           ||||||||||||||
Db      15 AACGAATTCCGGGT 28


RESULT   4
AA898689/c
LOCUS       AA898689      543 bp    mRNA            EST       12-APR-1998
DEFINITION  NCC3A9T7 Conidial Neurospora crassa cDNA clone NC3A9 3' end, mRNA
            sequence.
ACCESSION   AA898689
VERSION     AA898689.1  GI:3045122
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;
            Neurospora.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K.,
            Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
            Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
            and Natvig,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2282259.
            Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu
            Deposited in GSDB at the National Center for Genome Resources with
            accession GSDB:S:1147380
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Neurospora crassa"
                     /strain="74-OR23-IV A"
                     /db_xref="taxon:5141"
                     /clone="NC3A9"
                     /clone_lib="Conidial"
                     /sex="Mating type A"
                     /tissue_type="Conidia"
                     /dev_stage="germinating conidia"
                     /note="mRNA isolated from germinating conidia, grown in 1x
                     Vogel's, 2% sucrose for 4.5 hours.  cDNA directionally
                     cloned into pBluescript SK(-) using the Uni-ZAP XR vector
                     system (Stratagene, La Jolla, CA)."
BASE COUNT      108 a    163 c    141 g    131 t
ORIGIN


  Query Match             93.3%;  Score 14;  DB 39;  Length 543;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGT 14
           ||||||||||||||
Db     154 AACGAATTCCGGGT 141
```

```
RESULT   5
AQ574334/c
LOCUS       AQ574334      871 bp    DNA             GSS       02-JUN-1999
DEFINITION  nbxb0085G16f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0085G16f, genomic survey sequence.
ACCESSION   AQ574334
VERSION     AQ574334.1  GI:4974819
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 378.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0085G16f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                      of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep  coverage allows the isolation a particular
                     sequence with a probability  of 99.9 %. Two high density
                     filters, each containing 18,432 clones  (doubly spotted),
                     represent the whole library for colony screening."
BASE COUNT       258 a    167 c    165 g    281 t
ORIGIN


  Query Match             93.3%;  Score 14;  DB 108;  Length 871;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGT 14
           ||||||||||||||
Db     827 AACGAATTCCGGGT 814


RESULT   6
AV081419
LOCUS       AV081419      160 bp    mRNA            EST       25-JUN-1999
DEFINITION  AV081419 Mus musculus stomach C57BL/6J adult Mus musculus cDNA
            clone 2210418O05, mRNA sequence.
ACCESSION   AV081419
VERSION     AV081419.1  GI:5212867
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 160)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 16, 1998 this sequence version replaced gi:2961935.
            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .160
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2210418O05"
                     /clone_lib="Mus musculus stomach C57BL/6J adult"
                     /sex="male"
                     /tissue_type="stomach"
                     /dev_stage="adult"
BASE COUNT        44 a     23 c     47 g     46 t
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 59;  Length 160;
  Best Local Similarity   93.3%;  Pred. No. 2.3e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 AACGAATTCCGGGTG 15
           |||||||||| ||||
Db      10 AACGAATTCCAGGTG 24


RESULT   7
W05084
LOCUS       W05084        182 bp    mRNA            EST       23-APR-1996
DEFINITION  za79b12.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:298751 5' similar to gb:M13755 INTERFERON-INDUCED 17 KD
            PROTEIN (HUMAN);, mRNA sequence.
ACCESSION   W05084
VERSION     W05084.1  GI:1277806
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182)
  AUTHORS   Hill,R.,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```

```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 8, 1995 this sequence version replaced gi:800042.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 154.
FEATURES             Location/Qualifiers
     source          1. .182
                     /organism="Homo sapiens"
                     /db_xref="GDB:1243675"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:298751"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharma a) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT       40 a     55 c     54 g     31 t      2 others
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 25;  Length 182;
  Best Local Similarity   93.3%;  Pred. No. 2.3e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          ||||||||||| ||||
Db     98 AACGAATTCCAGGTG 112


RESULT   8
T28507
LOCUS       T28507         209 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST47206 Human Kidney Homo sapiens cDNA 3' end similar to
            interferon-induced protein, 15 kDa/17 kDa (HT:412), mRNA sequence.
ACCESSION   T28507
VERSION     T28507.1  GI:610605
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., , Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):102784"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Kidney"
                     /note="Organ: kidney"
BASE COUNT       44 a     67 c     62 g     35 t      1 others
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 20;  Length 209;
  Best Local Similarity   93.3%;  Pred. No. 2.4e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          ||||||||||| ||||
Db    100 AACGAATTCCAGGTG 114


RESULT   9
AW392780/c
LOCUS       AW392780       210 bp    mRNA            EST       04-FEB-2000
DEFINITION  CM3-ST0274-011299-049-d02 ST0274 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW392780
VERSION     AW392780.1  GI:6897439
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 210)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 9, 1999 this sequence version replaced gi:5866790.
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-ST0274-
            011299-049-d02&t3=1999-12-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 210.
```

```
FEATURES             Location/Qualifiers
     source          1. .210
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ST0274"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       53 a     25 c     54 g     78 t
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 80;  Length 210;
  Best Local Similarity   93.3%;  Pred. No. 2.4e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          ||||||||| |||||
Db    139 AACGAATTCAGGGTG 125


RESULT  10
AL036722/c
LOCUS       AL036722      234 bp    mRNA            EST       27-SEP-1999
DEFINITION  DKFZp564G1563_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
            DKFZp564G1563 5', mRNA sequence.
ACCESSION   AL036722
VERSION     AL036722.3  GI:5927877
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Duesterhoeft, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     On Jul 7, 1999 this sequence version replaced gi:5866276.
            Contact: Duesterhoeft A
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp564G1563"
                     /clone_lib="564 (synonym: hfbr2)"
                     /tissue_type="brain"
                     /dev_stage="fetal"
                     /lab_host="Xl-2blue"
                     /note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT       59 a     32 c     50 g     90 t      3 others
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 64;  Length 234;
  Best Local Similarity   93.3%;  Pred. No. 2.5e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          ||||||||| |||||
Db    196 AACGAATTCAGGGTG 182


RESULT  11
C45133/c
LOCUS       C45133        240 bp    mRNA            EST       18-OCT-1999
DEFINITION  C45133 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk376f1 5', mRNA sequence.
ACCESSION   C45133
VERSION     C45133.1  GI:2381370
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802362.
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk376f1"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       58 a     51 c     72 g     53 t      6 others
ORIGIN


  Query Match             89.3%;  Score 13.4;  DB 35;  Length 240;
  Best Local Similarity   93.3%;  Pred. No. 2.5e+02;
  Matches   14;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 AACGAATTCCGGGTG 15
          |||| ||||||||||
Db    167 AACGCATTCCGGGTG 153


RESULT  12
AA767764/c
LOCUS       AA767764      248 bp    mRNA            EST       08-FEB-1998
DEFINITION  ob37d05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1333545 3',
            mRNA sequence.
ACCESSION   AA767764
VERSION     AA767764.1  GI:2818779
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 248)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152154.
            Contact: Robert Strausberg, Ph.D.
```

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 809 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 240.

FEATURES Location/Qualifiers
source 1. .248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1333545"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 64 a 27 c 62 g 95 t
ORIGIN

Query Match 89.3%; Score 13.4; DB 38; Length 248;
Best Local Similarity 93.3%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 AACGAATTCCGGGTG 15
          ||||||||| |||||
Db    160 AACGAATTCAGGGTG 146
```

RESULT 13
H41052/c
LOCUS H41052 256 bp mRNA EST 16-AUG-1995
DEFINITION yp52c07.s1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:191052 3', mRNA sequence.
ACCESSION H41052
VERSION H41052.1 GI:917104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 256)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:803951.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1304
High quality sequence stops: 225
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1304 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 225.

FEATURES Location/Qualifiers
source 1. .256
/organism="Homo sapiens"
/db_xref="GDB:3847461"
/db_xref="taxon:9606"
/clone="IMAGE:191052"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 62 a 31 c 65 g 88 t 10 others
ORIGIN

Query Match 89.3%; Score 13.4; DB 23; Length 256;
Best Local Similarity 93.3%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 AACGAATTCCGGGTG 15
          ||||||||| |||||
Db    152 AACGAATTCAGGGTG 138
```

RESULT 14
N90615/c
LOCUS N90615 256 bp mRNA EST 03-APR-1996
DEFINITION zb42h08.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone IMAGE:306303 3', mRNA sequence.
ACCESSION N90615
VERSION N90615.1 GI:1443942
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 256)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785097.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 141.

FEATURES Location/Qualifiers
source 1. .256
/organism="Homo sapiens"
/db_xref="GDB:1249713"
/db_xref="taxon:9606"
/clone="IMAGE:306303"
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTT T-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 66 a 32 c 64 g 92 t 2 others
ORIGIN

Query Match 89.3%; Score 13.4; DB 25; Length 256;
Best Local Similarity 93.3%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACGAATTCCGGGTG 15
||||||||| |||||
Db 152 AACGAATTCAGGGTG 138

RESULT 15
W24444
LOCUS W24444 262 bp mRNA EST 20-AUG-1996
DEFINITION zb60b04.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:307951 5' similar to gb:M13755 INTERFERON-INDUCED 17 KD PROTEIN (HUMAN);, mRNA sequence.
ACCESSION W24444
VERSION W24444.1 GI:1301325
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 262)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 687 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 161.

FEATURES Location/Qualifiers
source 1. .262
/organism="Homo sapiens"
/db_xref="GDB:1251364"
/db_xref="taxon:9606"
/clone="IMAGE:307951"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHH19W."

BASE COUNT 49 a 82 c 78 g 44 t 9 others
ORIGIN

Query Match 89.3%; Score 13.4; DB 26; Length 262;
Best Local Similarity 93.3%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACGAATTCCGGGTG 15
||||||||| ||||
Db 92 AACGAATTCCAGGTG 106

Search completed: June 22, 2000, 17:43:53
Job time: 20103 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:15:33 ; Search time 40.17 Seconds
(without alignments)
2.948 Million cell updates/sec

Title: US-09-362-485-29
Perfect score: 26
Sequence: 1 NEFRV 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	246	1	P90613	Portion of VP3 of
2	26	100.0	276	1	Y06964	E. canis P30-4 pro
3	26	100.0	384	1	R15056	Capsid region of c
4	26	100.0	516	1	R77522	Cytochrome-P450 CS
5	26	100.0	516	1	R77511	Cytochrome-P450-1p
6	26	100.0	839	1	R15629	Capsid region of c
7	26	100.0	854	1	P50287	Sequence encoded b
8	26	100.0	993	1	P50116	Sequence of Hepati
9	26	100.0	993	1	P50231	Sequence encoded b
10	26	100.0	1077	1	W95559	A partial hepatiti
11	26	100.0	1091	1	R32426	Translated from 5'
12	26	100.0	2227	1	R05697	Attenuated hepatit
13	26	100.0	2227	1	P60066	Sequence of viral
14	26	100.0	2227	1	W34074	Hepatitis A virus
15	25	96.2	402	1	W61474	Comamonas sp. N-Ac
16	25	96.2	884	1	W59778	Amino acid sequenc
17	25	96.2	884	1	W71375	Yeast telomerase c
18	23	88.5	128	1	W55197	H. pylori ORF 01ce
19	23	88.5	141	1	W00966	GCN4-C/EBP fusion.
20	23	88.5	159	1	W74907	Human secreted pro
21	23	88.5	192	1	Y11046	H. pylori ORF 14ee
22	23	88.5	246	1	R64256	Mamestra brassicae
23	23	88.5	258	1	Y11045	H. pylori ORF 07ee
24	23	88.5	291	1	R53085	Transcription acti
25	23	88.5	329	1	W55718	H. pylori ORF 14cp
26	23	88.5	329	1	W71499	Helicobacter polyp
27	23	88.5	360	1	R62381	Human CCAAT/Enhanc
28	23	88.5	360	1	W06797	CCAAT/Enhancer bin
29	23	88.5	389	1	W44948	G-alpha s protein.
30	23	88.5	394	1	R94559	Human Gs alpha pro
31	23	88.5	394	1	W61179	Chimera consisting
32	23	88.5	394	1	W61180	Chimera consisting
33	23	88.5	394	1	W61181	Chimera consisting
34	23	88.5	394	1	W61182	Chimera consisting
35	23	88.5	394	1	W61183	Chimera consisting
36	23	88.5	394	1	W61184	Chimera consisting
37	23	88.5	394	1	W61185	Chimera consisting
38	23	88.5	394	1	W61186	Chimera consisting
39	23	88.5	394	1	W61187	Chimera consisting
40	23	88.5	394	1	W61188	Wild type G-alpha
41	23	88.5	412	1	W05261	Chromatin regulato
42	23	88.5	430	1	W85600	Hexosaminidase enz
43	23	88.5	492	1	W01560	CYB1 protein. New
44	23	88.5	568	1	W53451	Mouse gamma-glutam
45	23	88.5	578	1	W55591	H. pylori ORF 09cp

ALIGNMENTS

RESULT 1
P90613
ID P90613 standard; protein; 246 AA.
AC P90613;
DT 10-MAR-1993 (revised)
DT 31-OCT-1989 (first entry)
DE Portion of VP3 of hepatitis A virus strain HM175
KW Hepatitis A virus; VP1; HM175 strain; DNA; immunogen;
KW conjugate; vaccines; VP3.
OS Hepatitis A virus.
PN WO8906658-A.
PD 27-JUL-1989.
PF 13-JAN-1989;U00097.
PR 13-JAN-1988; US-143395.
PA (UYNC) Univ of N Carolina.
PI Lemon S M;
DR WPI; 89-233826/32.
DR N-PSDB; N90400.
PT Synthetic immunogenic peptide(s) corrsp. to hepat.itis A virus
PT - derived from VP3 struct.:al protein and used for vaccine prodn.
PT and diagnosis.
PS Disclosure; fig :; 29pp; English.
CC Portion of VP3 (see N90400) structural protein of
CC HM175 strain hepatitis A virus. This region corresponds to
CC an immunogenic peptide contg. Asp-70. It can form an immunogenic
CC conjugate with eg a portion of VP1 (contg. Ser-102; see N90399,
CC P90612). The immunogens can be used to produce and detect
CC antibodies, and in vaccines against HAV.
SQ Sequence 246 AA; 17 A; 7 C; 11 G; 22 T; 189 other;

Query Match 100.0%; Score 26; DB 1; Length 246;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
||||| |
Db 4 NEFRV 8

RESULT 2
Y06964
ID Y06964 standard; Protein; 276 AA.
AC Y06964;
DT 05-JUL-1999 (first entry)
DE E. canis P30-4 protein.
KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;
KW detection; dog.
OS Ehrlichia canis.
PN WO9913720-A1.
PD 25-MAR-1999.
PF 18-SEP-1998; U19600.
PR 19-SEP-1997; US-059353.
PA (OHIS) UNIV OHIO STATE.
PI Ohashi N, Rikihisa Y;
DR WPI; 99-254290/21.
DR N-PSDB; X34764.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PS Disclosure; Fig 24B; 55pp; English.
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chafeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in Y06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in Y06959-970. The proteins and genes are used
CC to detect E. chaffeensis in patients and E. canis in dogs.
SQ Sequence 276 AA;

Query Match 100.0%; Score 26; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 237 NEFRV 241

RESULT 3
R15056
ID R15056 standard; Protein; 384 AA.
AC R15056;
DT 17-MAR-1992 (first entry)
DE Capsid region of cyno-HAV isolate CY-55/JM-55.
KW Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.
OS Cynomolgus monkey hepatitis A virus, isolate CY-55/JM-55.
FH Key Location/Qualifiers
FT protein 1. .76
FT /label= VP0
FT protein 77. .169
FT /label= VP3
FT protein 170. .284
FT /label= VP1
FT cleavage_site 76. .77
FT cleavage_site 169. .170
FT active_site 146
FT active_site 270
PN US7678828-A.
PD 12-NOV-1991.
PF 03-APR-1991; 678828.
PR 03-APR-1991; US-678828.
PA (USSH) US DEPT HEALTH & HUMAN.
PI Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;
DR WPI; 91-376737/51.
DR N-PSDB; Q13697.
PT Hepatitis A virus isolates and DNA - used to prepare vaccines for
PT preventing hepatitis A virus infection.
PS Disclosure; Fig 3; 23pp; English.
CC The sequence was deduced from the nucleotide sequence obtd. by PCR
CC amplification of cyno-HAV viral RNA obtd. from the stool of a
CC cynomolgus macaque inocculated i.v. with a 10% liver homogennate
CC obtd. from a macaque imported from Indonesia that spontaneously
CC developed HAV infection. The sequence differs from the human HAV
CC isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.
CC USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val
CC pair at the VP3-VP1 cleavage site in the human isolate is replaced
CC by a Gln-Thr pair in the cyno-HAV.The other two cleavage sites are
CC the same. Two residues have been identified as part of the immuno-
CC dominant region (see feature table) and are different to those in
CC the same position in human HAV. The protein and peptides derived
CC from it can be used in the prepn. of vaccines for the prevention of
CC HAV infection.
CC See also R15629.
SQ Sequence 384 AA;

Query Match 100.0%; Score 26; DB 1; Length 384;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 80 NEFRV 84

RESULT 4
R77522
ID R77522 standard; Protein; 516 AA.
AC R77522;
DT 20-MAR-1996 (first entry)
DE Cytochrome-P450 CS.
KW Cytochrome P450; P450-monooxygenase; insecticide; housefly.
OS Musca domestica CS.
PN WO9530745-A1.
PD 16-NOV-1995.
PF 08-MAY-1995; U05758.
PR 10-MAY-1994; US-241388.
PA (CORR) CORNELL RES FOUND INC.
PI Scott JG, Tomita T;
DR WPI; 95-404112/51.
DR N-PSDB; T05526.
PT DNA encoding cytochrome P450-1pr - used for insect control,
PT bioremediation of insecticides or reducing crop sensitivity to
PT pesticides
PS Example 11; Page 39-40; 87pp; English.
CC The amino acid sequence of cytochrome P450-cs (R77522) isolated
CC from insecticide-susceptible adult housefly strain CS showed only
CC 2 differences from that of P450-1pr, a cytochrome capable of
CC metabolising insecticides, obtd. from housefly strain Learn-PyR.
CC Ile-220 and Ile-469 in P450-1pr are replaced by Phe and Met,
CC respectively, in CS.
SQ Sequence 516 AA;

Query Match 100.0%; Score 26; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 199 NEFRV 203

RESULT 5
R77511
ID R77511 standard; Protein; 516 AA.
AC R77511;
DT 20-MAR-1996 (first entry)
DE Cytochrome-P450-1pr.
KW Cytochrome P450-1pr; P450-monooxygenase; housefly; insecticide;
KW biological control; pesticide degradation; bioremediation;
KW transgenic plant; crop improvement; insect resistance.
OS Musca domestica Learn-PyR.
FH Key Location/Qualifiers
FT peptide 1. .19
FT /label= Sig_peptide
FT /note= "amino acids 1-19 constitute a
FT membrance-anchor signal"
FT protein 20. .516
FT /label= Mat_protein
PN WO9530745-A1.
PD 16-NOV-1995.
PF 08-MAY-1995; U05758.
PR 10-MAY-1994; US-241388.
PA (CORR) CORNELL RES FOUND INC.
PI Scott JG, Tomita T;
DR WPI; 95-404112/51.
DR N-PSDB; T05517.
PT DNA encoding cytochrome P450-1pr - used for insect control,
PT bioremediation of insecticides or reducing crop sensitivity to
PT pesticides
PS Claim 5; Page 44-46; 87pp; English.
CC Cytochrome P450-1pr (R77511) obtd. from adult housefly strain

CC Learn-PyR, selected with permethrin for 22 generations, is capable of
CC metabolising pyrethroids and activating organophosphate insecticides.
CC Recombinant P450-1pr, produced by expression of encoding cDNA (T05517),
CC can be used for insecticide bioremediation. Expression in e.g. a
CC baculovirus vector allows biological control of larval and adult insect
CC pests, while expression in transgenic plants reduces crop sensitivity
CC to pesticides.
SQ Sequence 516 AA;

Query Match 100.0%; Score 26; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 199 NEFRV 203

RESULT 6
R15629
ID R15629 standard; Protein; 839 AA.
AC R15629;
DT 17-MAR-1992 (first entry)
DE Capsid region of cyno-HAV isolate CY-145.
KW Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.
OS Cynomolgus monkey hepatitis A virus, isolate CY-145.
FH Key Location/Qualifiers
FT protein 1. .245
FT /label= VP0
FT protein 246. .491
FT /label= VP3
FT protein 492. .791
FT /label= VP1
FT protein 792
FT /label= P2
FT /note= "incomplete"
FT cleavage_site 245. .246
FT cleavage_site 491. .492
FT cleavage_site 791. .792
FT active_site 315
FT active_site 593
PN US7678828-A.
PD 12-NOV-1991.
PF 03-APR-1991; 678828.
PR 03-APR-1991; US-678828.
PA (USSH) US DEPT HEALTH & HUMAN.
PI Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;
DR WPI; 91-376737/51.
DR N-PSDB; Q15180.
PT Hepatitis A virus isolates and DNA - used to prepare vaccines for
PT preventing hepatitis A virus infection.
PS Disclosure; Fig 3; 23pp; English.
CC The sequence was deduced from the nucleotide sequence obtd. by PCR
CC amplification of cyno-HAV viral RNA obtd. from the stool of a
CC cynomolgus monkey with serologically and histologically confirmed
CC spontaneous hepatitis A. The sequence differs from the human HAV
CC isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.
CC USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val
CC pair at the VP3-VP1 cleavage site in the human isolate is replaced
CC by a Gln-Thr pair in the cyno-HAV.The other two cleavage sites are
CC the same. Two residues have been identified as part of the immuno-
CC dominant region (see feature table) and are different to those in
CC the same position in human HAV. The protein and peptides derived
CC from it can be used in the prepn. of vaccines for the prevention of
CC HAV infection.
CC See also R15056.
SQ Sequence 839 AA;

Query Match 100.0%; Score 26; DB 1; Length 839;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 249 NEFRV 253

RESULT 7
P50287
ID P50287 standard; Protein; 854 AA.
AC P50287;
DT 30-NOV-1991 (first entry)
DE Sequence encoded by hepatitis A virus (HAV) cDNA from near the
DE genome 5' terminus to the end of the area corresponding to the
DE capsid protein region of poliovirus RNA.
KW Hepatitis A virus assay; antigen; antibody.
OS Hepatitis A virus.
PN WO8501517-A.
PD 11-APR-1985.
PF 27-SEP-1984; U01552.
PR 30-SEP-1983; US-537911.
PA (MASI) MASSACHUSETTS INST TECH.
PI Ticechurst JR, Baltimore D, Feinstone SM, Purcell RH,
PI Racaniello VR;
DR WPI; 85-098846/16.
DR N-PSDB; N50330.
PT New hepatitis A virus CDNA - useful in assays for the virus and
PT for prodn. of the viral antigen and antibodies to it
PS Example; Fig 7; 60pp; English.
CC The inventors claim HAV cDNA and a method for producing it, whereby
CC large amts. can be obtd. economically. The cDNA is useful in the
CC assay for detection of HAV quickly and easily and with high
CC sensitivity and specificity. The HAV cDNA is also used in the prodn.
CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.
SQ Sequence 854 AA;

Query Match 100.0%; Score 26; DB 1; Length 854;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 249 NEFRV 253

RESULT 8
P50116
ID P50116 standard; Protein; 993 AA.
AC P50116;
DT 30-SEP-1991 (first entry)
DE Sequence of Hepatitis A virus (HAV) immunogenic peptides
DE VP-1, VP-2, VP-3 and VP-4.
KW Antigenic protein; immunogen; vaccine.
OS Hepatitis A virus (strain CR326).
PN EP-154587-A.
PD 11-SEP-1985.
PF 27-FEB-1985; 400369.
PR 02-MAR-1984; US-585818.
PA (MERI) MERCK & CO INC.
PI Linemeyer DL, Menke JG, Reuben RG, Mitra SW;
DR WPI; 85-224964/37.
DR N-PSDB; N50139.
PT New nucleotide sequences coding for hepatitis A virus antigens -
PT useful for eliciting normal immune response and in vaccines for
PT protecting against the virus
PS Example; Page 11-17; 32pp; English.
CC Within the sequence in N50139 is encoded the information necessary
CC to make the antigenic proteins of HAV. The sequences encoding for
CC the structural proteins begin at base 403. The key sub-unit
CC sequences within VP-1, designated Sequences I,II,III,IV, and V,
CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other
CC nucleotide sequences which are valuable as encoding antigenic
CC proteins are the sequences from base 1749 to base 2722; from base

```
CC   1487 to base 2980 and from base 1644 to base 2722. The sequence from
CC   base 1749 to base 2722 is esp. valuable as a vector for producing
CC   antigen protein. Sequences II-V are claimed. X in P50116 denotes the
CC   translation of a stop codon.
SQ   Sequence   993 AA;

  Query Match             100.0%;  Score 26;  DB 1;  Length 993;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     385 NEFRV 389


RESULT   9
P50231
ID   P50231 standard; Protein; 993 AA.
AC   P50231;
DT   28-NOV-1991  (first entry)
DE   Sequence encoded by partial sequence of hepatitis A virus (HAV),
DE   including surface protein (VP-1).
KW   Hepatitis A virus vaccine; immunisation; monoclonal antibody;
KW   diagnostic assay.
OS   Hepatitis A virus.
FH   Key             Location/Qualifiers
FT   protein         628. .993
FT                   /note= "claimed; X denotes translated stop codons
FT                   and unspecified triplets"
PN   EP-138704-A.
PD   24-APR-1985.
PF   09-OCT-1984; 402025.
PR   14-OCT-1983; US-541836.
PR   02-MAR-1984; US-585942.
PA   (MERI ) MERCK & CO INC.
PI   Hughes JV, Scolnick EM, Tomassini JE;
DR   WPI; 85-100818/17.
DR   N-PSDB; N50274.
PT   New hepatitis A virus surface protein - useful for binding to
PT   neutralising antibodies to the virus
PS   Disclosure; Page 17-23; 49pp; English.
CC   VP1 is isolated by solubilisation of the intact virus in an aq.
CC   anionic surfactant and a reducing agent. The viral proteins are sepd.
CC   and the protein of molecular wt. 33000 daltons is sepd.
SQ   Sequence   993 AA;

  Query Match             100.0%;  Score 26;  DB 1;  Length 993;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     385 NEFRV 389


RESULT   10
W95559
ID   W95559 standard; Protein; 1077 AA.
AC   W95559;
DT   28-APR-1999  (first entry)
DE   A partial hepatitis A virus (HAV) protein.
KW   Hepatitis A virus protein; HAV; P2 region;
KW   cell-culture-adapted HAV strain; infection; accelerated growth.
OS   Hepatitis A virus.
PN   US5849562-A.
PD   15-DEC-1998.
PF   06-JUN-1995; 468926.
PR   06-NOV-1991; US-788262.
PR   30-SEP-1983; US-537911.
PR   27-SEP-1984; US-654942.
PR   06-OCT-1988; US-256135.
PR   06-JUN-1995; US-468926.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Emerson SU, Purcell RH;
DR   WPI; 99-094412/08.
DR   N-PSDB; X01006.
PT   Chimeric hepatitis A virus strains - with P2 region from
PT   cell-culture-adapted strain in wild-type genome
PS   Disclosure; Fig 7A-L; 36pp; English.
CC   The present sequence represents a partial hepatitis A virus (HAV)
CC   protein. The specification describes a DNA construct consisting
CC   of a wild-type HAV genome in which the P2 region is replaced by the
CC   P2 region from a cell-culture-adapted HAV strain. The construct is
CC   used to demonstrate that mutations in the P2 region of a
CC   cell-culture-adapted HAV strain are sufficient for establishment of
CC   infection and accelerated growth in cell culture.
SQ   Sequence   1077 AA;

  Query Match             100.0%;  Score 26;  DB 1;  Length 1077;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     472 NEFRV 476


RESULT   11
R32426
ID   R32426 standard; Protein; 1091 AA.
AC   R32426;
DT   10-JUN-1993  (first entry)
DE   Translated from 5' region of Hepatitis A Virus genomic clone.
KW   HAV HM-175; chronic liver disease; picornavirus.
OS   Hepatitis A Virus.
FH   Key             Location/Qualifiers
FT   region          238. .1091
FT                   /label= ORF
FT                   /note= "second putative initiation codon at
FT                   position 240"
FT   region          1. .711
FT                   /note= "X's correspond to nonsense codons,
FT                   i.e. this region is not an ORF"
PN   US7788262-A.
PD   15-DEC-1992.
PF   30-SEP-1983; 536911.
PR   30-SEP-1983; US-536911.
PR   27-SEP-1984; US-654942.
PR   06-OCT-1988; US-256135.
PR   06-NOV-1991; US-788262.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Baltimore D,  Feinstone SM;
PI   Purcell RH, Racaniello VR,  Ticehurst JR;
DR   WPI; 93-067429/08.
DR   N-PSDB; Q36934.
PT   Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.
PT   of antigen and antibodies
PS   Disclosure; Fig 7; 65pp; English.
CC   HAV virion RNA was extracted from the livers of marmosets which had
CC   been inoculated with HAV (the HAV had previously been passaged twice
CC   in marmosets). The RNA was used to prepare ds cDNA clones by
CC   standard methods. Clones contg. inserts which hybridised to RNA from
CC   HAV-infected African Green Monkey Kidney cells were selected for
CC   further analysis. A 7.4kb restriction map (about 99% of the HAV
CC   genome) was constructed from 5 overlapping inserts. The sequence of
CC   the first 3.3kb (approx.) from the 5'-terminus was determined. An
CC   amino acid sequence was decoded from the entire clone and an open
CC   reading frame was identified starting at position 238. A comparison
CC   of the predicted HAV amino acid sequences with the known capsid
CC   protein sequences of other picornaviruses (poliovirus, foot and
CC   mouth disease virus and encephalomyelitis virus) revealed areas of
CC   local homology.
SQ   Sequence   1091 AA;
```

```
  Query Match             100.0%;  Score 26;  DB 1;  Length 1091;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     486 NEFRV 490


RESULT  12
R05697
ID   R05697 standard; protein; 2227 AA.
AC   R05697;
DT   15-AUG-1990  (first entry)
DE   Attenuated hepatitis A virus
KW   Hepatitis A virus; vaccine; attenuated.
OS   Hepatitis A virus, strain HM-175.
FH   Key             Location/Qualifiers
FT   region          1. .23
FT                   /label=VP4 - 1A
FT   region          24. .245
FT                   /label=VP2 - 1B
FT   region          246. .491
FT                   /label=VP3 - 1C
FT   region          492. .791
FT                   /label=VP1 - 1D
FT   region          792. .980
FT                   /label=2A
FT   region          981. .1087
FT                   /label=2B
FT   region          1088. .1422
FT                   /label=2C
FT   region          1423. .1496
FT                   /label=3A
FT   region          1497. .1519
FT                   /label=3B - VPg
FT   region          1520. .1738
FT                   /label=3C
FT   region          1739. .2227
FT                   /label=3D
PN   US4894228-A.
PD   16-JAN-1990.
PF   12-JUL-1988; 217824.
PR   12-JUL-1988; US-217824.
PR   12-JUL-1988; US-652967.
PA   (USSH) US Dept Health & Human.
PI   Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstone SM,
PI   Daemer RJ, Gust ID;
DR   WPI; 90-075557/10.
DR   N-PSDB; Q03512.
PT   Vaccine against hepatitis A virus  infection -  comprises novel
PT   attenuated hepatitis A virus strain.
PS   Claim 1; Fig 1; 18pp; English.
CC   The attenuated HAV is useful for inducing protective immunity against
CC   HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by
CC   several nucleotide changes distributed throughout the genome, is
CC   attenuated for chimpanzees, elicits serum neutralising antibodies, and is
CC   suitable for use as an HAV vaccine. It is noted that not all the changes
CC   are necessary for attenuation and use as a vaccine.
SQ   Sequence   2227 AA;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  13
P60066
ID   P60066 standard; Protein; 2227 AA.
AC   P60066;
DT   26-JUN-1991  (first entry)
DE   Sequence of viral L434 polypeptide encoded by the complete
DE   nucleotide sequence of the HAV genome.
KW   Diagnosis; vaccine; passive immunotherapy.
OS   Hepatitis A virus.
FH   Key             Location/Qualifiers
FT   region          1. .245
FT                   /label= P1.1A
FT   region          246. .491
FT                   /label= 1B
FT   region          492. .836
FT                   /label= 1C
FT   region          837. .980
FT                   /label= P2.2A
FT   region          981. .1076
FT                   /label= 2B
FT   region          1077. .1422
FT                   /label= 2C
FT   region          1423. .1484
FT                   /label= P3.3A
FT   region          1485. .1507
FT                   /label= 3B
FT   region          1508. .1678
FT                   /label= 3C
FT   region          1679. .2227
FT                   /label= 3D
PN   EP-199480-A.
PD   29-OCT-1986.
PF   03-APR-1986; 302465.
PR   03-APR-1985; US-719329.
PA   (CHIR-) CHIRON CORP.
PI   Dina D, Potter SJ, Vannest GA, Caput D;
DR   WPI; 86-286213/44.
DR   N-PSDB; N60080.
PT   Hepatitis A virus nucleotide sequence and polypeptide - and use
PT   in prodn. of vaccines and diagnostic probes
PS   Claim 5; Fig 1; 18pp; English.
CC   N60080 and oligonucleotide fragments are useful in detection of
CC   hepatitis A virus; transformed hosts may be used for expression of
CC   polypeptides and fragments useful in vaccines without risk of
CC   infection by the virus or in prodn. of particles which are capable
CC   of inducing immmunocompetent B cells for passive immunotherapy. Pref.
CC   epitope is derived from AAs 445-657 or 792-848 of the HAV
CC   polypeptide sequence (P60066).
SQ   Sequence   2227 AA;


  Query Match             100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  14
W34074
ID   W34074 standard; Protein; 2227 AA.
AC   W34074;
DT   27-APR-1998  (first entry)
DE   Hepatitis A virus HM-175 protein sequence.
KW   HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;
KW   infection; vaccine.
OS   Hepatitis A virus HM-175.
FH   Key             Location/Qualifiers
FT   Protein         1. .23
FT                   /label= VP4
FT   Protein         24. .245
```

```
FT                   /label= VP2
FT   Protein         246. .491
FT                   /label= VP3
FT   Protein         492. .791
FT                   /label= VP1
FT   Protein         792. .980
FT                   /label= 2A
FT   Protein         981. .1087
FT                   /label= 2B
FT   Protein         1088. .1422
FT                   /label= 2C
FT   Protein         1423. .1496
FT                   /label= 3A
FT   Protein         1497. .1519
FT                   /label= 3B
FT   Protein         1520. .1738
FT                   /label= 3C
FT   Protein         1739. .2227
FT                   /label= 3D
PN   WO9740166-A2.
PD   30-OCT-1997.
PF   18-APR-1997; U06506.
PR   19-APR-1996; US-015642.
PA   (USSH ) US SEC DEPT HEALTH.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Emerson SU, Purcell RH, Raychaudhuri G;
DR   WPI; 97-535850/49.
DR   N-PSDB; T93023.
PT   Human attenuated HAV genome containing simian HAV 2C gene - useful
PT   as vaccines against HAV infection
PS   Disclosure; Fig 13A-D; 66pp; English.
CC   This protein sequence is encoded by the human hepatitis A virus
CC   (HAV) HM-175 wild-type genome (see T93023). Attenuated strain
CC   HAV/7 is obtained by passage of HM-175 in African Green Monkey
CC   kidney cells. A claimed DNA construct (I) comprises a genome of
CC   HAV, where the genome is a human attenuated HAV genome in which a
CC   region of the 2C gene has been replaced by a corresponding region
CC   from a 2C gene of a simian AGM-27 HAV genome (see T93024). The
CC   region of the 2C gene from AGM-27 contained in the construct
CC   preferably encodes amino acids 120-328 of the 2C protein, amino
CC   acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA
CC   transcript of (I); (2) a cell transfected with (I) or the RNA
CC   transcript of (1); (3) a HAV genome as above; (4) antibodies to the
CC   HAV of (3); and (5) a host cell containing the HAV of (3). (I) or
CC   its RNA transcript, can be used as a vaccine for preventing HAV in
CC   a mammal. (I) or the RNA transcript can also be used to stimulate
CC   the production of protective antibodies in the mammal.
SQ   Sequence   2227 AA;

  Query Match            100.0%;  Score 26;  DB 1;  Length 2227;
  Best Local Similarity  100.0%;  Pred. No. 4.7e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT  15
W61474
ID   W61474 standard; Protein; 402 AA.
AC   W61474;
DT   12-OCT-1998  (first entry)
DE   Comamonas sp. N-Acetylphosphinothricin deactylase protein.
KW   N-acetylphosphinothricin deacetylase; Stenotrophomonas; Comamonas;
KW   sterile; plant; L-phosphinothricin.
OS   Comamonas acidovorans.
PN   DE19652284-A1.
PD   18-JUN-1998.
PF   16-DEC-1996; 052284.
PR   16-DEC-1996; DE-052284.
PA   (AGRE ) HOECHST-SCHERING AGREVO GMBH.
PI   Bartsch K, Broer I, Kriete G, Puehler A;
DR   WPI; 98-334301/30.
DR   N-PSDB; V45110.
PT   DNA encoding N-acetyl:phosphinothricin deacetylase protein - useful
PT   for producing recombinant protein and male- or female-sterile plants
PS   Claim 2(d); Page 17-19; 24pp; German.
CC   The N-acetylphosphinothricin deacetylase gene can be isolated from
CC   both Stenotrophomonas sp. and Comamonas sp. . Plants can be engineered to
CC   contain the deacetylase gene, specific expression of this gene results in
CC   the deactivation of specific parts of the plant. This can be used for
CC   example to produce male- or female-sterile plants. The gene can also be
CC   used to produce recombinant N-acetylphosphinothricin deacetylase for use
CC   in the stereoselective preparation of L-phosphinothricin.
SQ   Sequence   402 AA;

  Query Match            96.2%;  Score 25;  DB 1;  Length 402;
  Best Local Similarity  80.0%;  Pred. No. 1.5e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           ||||:
Db     188 NEFRI 192


Search completed: June 22, 2000, 19:15:35
Job time: 5499 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:27:52 ; Search time 32.22 Seconds
(without alignments)
2.240 Million cell updates/sec

Title: US-09-362-485-29
Perfect score: 26
Sequence: 1 NEFRV 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	162	1	US-08-087-016-4	Sequence 4, Appli
2	26	100.0	517	1	US-08-457-274A-2	Sequence 2, Appli
3	26	100.0	517	1	US-08-457-274A-28	Sequence 28, Appl
4	26	100.0	517	4	PCT-US95-05758-2	Sequence 2, Appli
5	26	100.0	517	4	PCT-US95-05758-28	Sequence 28, Appl
6	26	100.0	839	1	US-08-087-016-2	Sequence 2, Appli
7	26	100.0	1091	5	5516630-2	Patent No. 5516630
8	23	88.5	31	2	US-08-690-011A-56	Sequence 56, Appl
9	23	88.5	31	2	US-08-690-011A-58	Sequence 58, Appl
10	23	88.5	141	2	US-08-690-011A-43	Sequence 43, Appl
11	23	88.5	207	2	US-08-808-550-30	Sequence 30, Appl
12	23	88.5	360	1	US-08-205-506A-2	Sequence 2, Appli
13	23	88.5	360	4	PCT-US94-02389-2	Sequence 2, Appli
14	23	88.5	380	3	US-08-307-896-1	Sequence 1, Appli
15	23	88.5	394	4	PCT-US95-11808-1	Sequence 1, Appli
16	23	88.5	492	1	US-08-463-090B-10	Sequence 10, Appl
17	23	88.5	587	1	US-07-955-905A-23	Sequence 23, Appl
18	23	88.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
19	23	88.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
20	23	88.5	1507	5	5268270-2	Patent No. 5268270
21	22	84.6	31	1	US-08-614-935-55	Sequence 55, Appl
22	22	84.6	105	1	US-08-081-539-73	Sequence 73, Appl
23	22	84.6	105	1	US-08-466-647-73	Sequence 73, Appl
24	22	84.6	111	1	US-08-411-795B-65	Sequence 65, Appl
25	22	84.6	111	1	US-08-411-795B-66	Sequence 66, Appl
26	22	84.6	111	1	US-08-411-795B-67	Sequence 67, Appl
27	22	84.6	111	1	US-08-411-795B-68	Sequence 68, Appl
28	22	84.6	111	1	US-08-411-795B-69	Sequence 69, Appl
29	22	84.6	111	1	US-08-411-795B-70	Sequence 70, Appl
30	22	84.6	111	1	US-08-411-795B-71	Sequence 71, Appl
31	22	84.6	111	1	US-08-411-795B-74	Sequence 74, Appl
32	22	84.6	111	1	US-08-411-795B-75	Sequence 75, Appl
33	22	84.6	111	1	US-08-411-795B-79	Sequence 79, Appl
34	22	84.6	111	1	US-08-411-795B-80	Sequence 80, Appl
35	22	84.6	111	1	US-08-411-795B-81	Sequence 81, Appl
36	22	84.6	111	1	US-08-411-795B-407	Sequence 407, App
37	22	84.6	111	1	US-08-411-796-23	Sequence 23, Appl
38	22	84.6	111	1	US-08-411-796-65	Sequence 65, Appl
39	22	84.6	111	1	US-08-411-796-217	Sequence 217, App
40	22	84.6	111	1	US-08-411-796-218	Sequence 218, App
41	22	84.6	111	2	US-08-469-319A-65	Sequence 65, Appl
42	22	84.6	111	2	US-08-469-319A-66	Sequence 66, Appl
43	22	84.6	111	2	US-08-469-319A-67	Sequence 67, Appl
44	22	84.6	111	2	US-08-469-319A-68	Sequence 68, Appl
45	22	84.6	111	2	US-08-469-319A-69	Sequence 69, Appl

ALIGNMENTS

```
RESULT   1
US-08-087-016-4
; Sequence 4, Application US/08087016
; Patent No. 5430135
;  GENERAL INFORMATION:
;    APPLICANT:  NAINAN, OMANA V.
;    APPLICANT:  MARGOLIS, HAROLD S.
;    APPLICANT:  ROBERTSON, BETTY H.
;    APPLICANT:  BRINTON, MARGO H.
;    APPLICANT:  EBERT, JAMES W.
;    TITLE OF INVENTION:  HEPATITIS A VIRUS VACCINE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1615 L Street N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,016
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/678,828
;      FILING DATE:  03-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26,581
;      REFERENCE/DOCKET NUMBER:  WTS/5683/83834/SRL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  162 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-087-016-4


  Query Match           100.0%;  Score 26;  DB 1;  Length 162;
  Best Local Similarity  100.0%;  Pred. No. 19;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 NEFRV 5
          |||||
Db     80 NEFRV 84


RESULT   2
US-08-457-274A-2
; Sequence 2, Application US/08457274A
; Patent No. 5734086
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Jeffrey G.
;    APPLICANT:  Tomita, Takashi
;    TITLE OF INVENTION:  Cytochrome P450lpr Gene and Its Uses
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,274A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/240 (D-1519)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  517 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-457-274A-2


  Query Match           100.0%;  Score 26;  DB 1;  Length 517;
  Best Local Similarity  100.0%;  Pred. No. 63;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NEFRV 5
          |||||
Db    199 NEFRV 203


RESULT   3
US-08-457-274A-28
; Sequence 28, Application US/08457274A
; Patent No. 5734086
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Jeffrey G.
;    APPLICANT:  Tomita, Takashi
;    TITLE OF INVENTION:  Cytochrome P450lpr Gene and Its Uses
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,274A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/240 (D-1519)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  517 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Musca domestica
;      STRAIN:  CS
;      DEVELOPMENTAL STAGE:  Adult
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  Chromosome 1
US-08-457-274A-28


  Query Match           100.0%;  Score 26;  DB 1;  Length 517;
  Best Local Similarity  100.0%;  Pred. No. 63;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NEFRV 5
          |||||
Db    199 NEFRV 203


RESULT   4
PCT-US95-05758-2
; Sequence 2, Application PC/TUS9505758
;  GENERAL INFORMATION:
;    APPLICANT:  Cornell Research Foundation, Inc.
;    TITLE OF INVENTION:  Cytochrome P450lpr Gene and Its
;    TITLE OF INVENTION:  Uses
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05758
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/241 (D-1519)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  517 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-05758-2


  Query Match             100.0%;  Score 26;  DB 4;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     199 NEFRV 203


RESULT   5
PCT-US95-05758-28
; Sequence 28, Application PC/TUS9505758
;  GENERAL INFORMATION:
;    APPLICANT:  Cornell Research Foundation, Inc.
;    TITLE OF INVENTION:  Cytochrome P450lpr Gene and Its
;    TITLE OF INVENTION:  Uses
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05758
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/241 (D-1519)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  517 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Musca domestica
;      STRAIN:  CS
;      DEVELOPMENTAL STAGE:  Adult
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  Chromosome 1
PCT-US95-05758-28


  Query Match             100.0%;  Score 26;  DB 4;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     199 NEFRV 203


RESULT   6
US-08-087-016-2
; Sequence 2, Application US/08087016
; Patent No. 5430135
;  GENERAL INFORMATION:
;    APPLICANT:  NAINAN, OMANA V.
;    APPLICANT:  MARGOLIS, HAROLD S.
;    APPLICANT:  ROBERTSON, BETTY H.
;    APPLICANT:  BRINTON, MARGO H.
;    APPLICANT:  EBERT, JAMES W.
;    TITLE OF INVENTION:  HEPATITIS A VIRUS VACCINE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN, DARBY & CUSHMAN
;      STREET:  1615 L Street N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,016
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/578,828
;      FILING DATE:  03-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26,581
;      REFERENCE/DOCKET NUMBER:  WTS/5683/83834/SRL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  839 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-087-016-2


  Query Match             100.0%;  Score 26;  DB 1;  Length 839;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     249 NEFRV 253


RESULT   7
5516630-2
;Patent No. 5516630
;    APPLICANT: TICEHURST, JOHN R.;BALTIMORE, DAVID;FEINSTONE,
;STEPHEN M.;PURCELL, ROBERT H.;RACANIELLO, VINCENT R.;
;BAROUDY, BAHIGE M.
;    TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/788,262
;      FILING DATE: 06-NOV-1991
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER: 256,135
;      FILING DATE: 06-OCT-1988
;      APPLICATION NUMBER: 654,942
;      FILING DATE: 27-SEP-1984
;      APPLICATION NUMBER: 537,911
;      FILING DATE: 30-SEP-1983
;SEQ ID NO:2:
;      LENGTH: 1091
5516630-2


  Query Match             100.0%;  Score 26;  DB 5;  Length 1091;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |||||
Db     486 NEFRV 490


RESULT   8
US-08-690-011A-56
; Sequence 56, Application US/08690011A
; Patent No. 5942433
;  GENERAL INFORMATION:
;    APPLICANT:  VINSON, Charles R.
;    APPLICANT:  KRYLOV, Dmitry
;    TITLE OF INVENTION:  EXTENSION OF A PROTEIN-PROTEIN
;    TITLE OF INVENTION:  INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
;    TITLE OF INVENTION:  CELLULAR PROTEIN
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,011A
;      FILING DATE:  31-JUL-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/001,654
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/018,496
;      FILING DATE:  29-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4199US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-690-011A-56


  Query Match             88.5%;  Score 23;  DB 2;  Length 31;
  Best Local Similarity   80.0%;  Pred. No. 17;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           ||:||
Db       2 NEYRV 6


RESULT   9
US-08-690-011A-58
; Sequence 58, Application US/08690011A
; Patent No. 5942433
;  GENERAL INFORMATION:
;    APPLICANT:  VINSON, Charles R.
;    APPLICANT:  KRYLOV, Dmitry
;    TITLE OF INVENTION:  EXTENSION OF A PROTEIN-PROTEIN
;    TITLE OF INVENTION:  INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
;    TITLE OF INVENTION:  CELLULAR PROTEIN
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 Park Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10154-0053
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,011A
;      FILING DATE:  31-JUL-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/001,654
;      FILING DATE:  31-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/018,496
;      FILING DATE:  29-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Serunian, Leslie A.
;      REGISTRATION NUMBER:  35,353
;      REFERENCE/DOCKET NUMBER:  2026-4199US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-690-011A-58


  Query Match             88.5%;  Score 23;  DB 2;  Length 31;
  Best Local Similarity   80.0%;  Pred. No. 17;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           ||:||
Db       2 NEYRV 6


RESULT  10
US-08-690-011A-43
; Sequence 43, Application US/08690011A
; Patent No. 5942433
;  GENERAL INFORMATION:
;    APPLICANT:  VINSON, Charles R.
;    APPLICANT:  KRYLOV, Dmitry
;    TITLE OF INVENTION:  EXTENSION OF A PROTEIN-PROTEIN
;    TITLE OF INVENTION:  INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
;    TITLE OF INVENTION:  CELLULAR PROTEIN
```

; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,011A
; FILING DATE: 31-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,654
; FILING DATE: 31-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/018,496
; FILING DATE: 29-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4199US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-690-011A-43

Query Match 88.5%; Score 23; DB 2; Length 141;
Best Local Similarity 80.0%; Pred. No. 83;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
||:||
Db 66 NEYRV 70

RESULT 11
US-08-808-550-30
; Sequence 30, Application US/08808550
; Patent No. 5871992
; GENERAL INFORMATION:
; APPLICANT: Teebor, George W.
; APPLICANT: Hilbert, Timothy P.
; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,550
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-001 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Haemophilus influenzae
US-08-808-550-30

Query Match 88.5%; Score 23; DB 2; Length 207;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
||:|
Db 169 NEFKV 173

RESULT 12
US-08-205-506A-2
; Sequence 2, Application US/08205506A
; Patent No. 5545563
; GENERAL INFORMATION:
; APPLICANT: Darlington, Gretchen J.
; APPLICANT: Wilson, Deborah R.
; APPLICANT: Wilde, Margaret
; TITLE OF INVENTION: THE HUMAN C/EBP GENE AND VECTORS FOR
; TITLE OF INVENTION: ITS EXPRESSION
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 801 Pennsylvania Avenue, NW
; CITY: Washington, D.C.
; STATE:
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/205,506A
; FILING DATE: March 4, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Davis, Peter J.
; REGISTRATION NUMBER: 36,119
; REFERENCE/DOCKET NUMBER: 311.027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 662-0200
; TELEFAX: (202) 662-4643
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-205-506A-2


  Query Match             88.5%;  Score 23;  DB 1;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           ||:||
Db     285 NEYRV 289


RESULT  13
PCT-US94-02389-2
; Sequence 2, Application PC/TUS9402389
;  GENERAL INFORMATION:
;    APPLICANT:  Darlington, Gretchen J.
;    APPLICANT:  Wilson, Deborah R.
;    APPLICANT:  Wilde, Margaret
;    TITLE OF INVENTION:  THE HUMAN C/EBP GENE AND VECTORS FOR ITS
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howrey & Simon
;      STREET:  1299 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/02389
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Auerbach, Jeffrey I.
;      REGISTRATION NUMBER:  32,680
;      REFERENCE/DOCKET NUMBER:  225-107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 383-7451
;      TELEFAX:  (202) 383-6610
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  360 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
PCT-US94-02389-2


  Query Match             88.5%;  Score 23;  DB 4;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           ||:||
Db     285 NEYRV 289


RESULT  14
US-08-307-896-1
; Sequence 1, Application US/08307896C
; Patent No. 6034071
; GENERAL INFORMATION:
;  APPLICANT:  Iyengar, Srinivas Ravi
;  TITLE OF INVENTION:  MUTANT ACTIVATED GSALPHA AND ADENYLYL
;  TITLE OF INVENTION:  CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;  FILE REFERENCE:  29770
;  CURRENT APPLICATION NUMBER:  US/08/307,896C
;  CURRENT FILING DATE:  1994-09-16
;  NUMBER OF SEQ ID NOS:  9
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH:  380
;   TYPE:  PRT
;   ORGANISM:  Homo sapiens
US-08-307-896-1


  Query Match             88.5%;  Score 23;  DB 3;  Length 380;
  Best Local Similarity   80.0%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 NEFRV 5
           |:|||
Db     110 NQFRV 114


RESULT  15
PCT-US95-11808-1
; Sequence 1, Application PC/TUS9511808
;  GENERAL INFORMATION:
;    APPLICANT:  Iyengar, Srinivas Ravi V.
;    TITLE OF INVENTION:  MUTANT ACTIVATED GsALPHA AND
;    TITLE OF INVENTION:  ADENYLYL
;    TITLE OF INVENTION:  CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue and
;      ADDRESSEE:  Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11808
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/307,896
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S.
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  29970 165/28755
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 408-2500
;      TELEFAX:  (212) 765-2519
;      TELEX:  650 6111063
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  394 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  GsALPHA
PCT-US95-11808-1
```

Query Match 88.5%; Score 23; DB 4; Length 394;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NEFRV 5
           |:|||
Db     124 NQFRV 128
```

Search completed: June 22, 2000, 19:27:53
Job time: 5324 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2000, 19:28:56 ; Search time 51.21 Seconds
(without alignments)
5.724 Million cell updates/sec

Title: US-09-362-485-29
Perfect score: 26
Sequence: 1 NEFRV 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	100.0	74	2	H70070	hypothetical prote
2	26	100.0	156	2	T28997	hypothetical prote
3	26	100.0	158	2	S39770	seminal vesicle au
4	26	100.0	267	2	S28670	avrPph3 protein -
5	26	100.0	300	2	C70930	hypothetical prote
6	26	100.0	343	2	T27185	hypothetical prote
7	26	100.0	369	2	B71624	hypothetical prote
8	26	100.0	371	1	A43830	alanine dehydrogen
9	26	100.0	371	2	T36876	probable L-alanine
10	26	100.0	384	2	PQ0223	polyprotein - simi
11	26	100.0	720	2	T15756	hypothetical prote
12	26	100.0	839	1	GNNYS2	genome polyprotein
13	26	100.0	852	1	GNNYHA	genome polyprotein
14	26	100.0	997	2	S63064	probable membrane
15	26	100.0	1160	3	T31688	Ca2+-transporting
16	26	100.0	1358	2	A03905	genome polyprotein
17	26	100.0	2227	1	GNNYHM	genome polyprotein
18	26	100.0	2227	1	GNNYHR	genome polyprotein
19	26	100.0	2227	1	GNNYMK	genome polyprotein
20	26	100.0	2227	1	GNNYHB	genome polyprotein
21	26	100.0	2230	1	GNNYSA	genome polyprotein
22	25	96.2	309	2	S46684	killer toxin insen
23	25	96.2	343	2	T29819	hypothetical prote
24	25	96.2	573	2	S44605	C02F5.3 protein -
25	25	96.2	605	2	S06398	alpha-globulin typ
26	25	96.2	884	2	S53396	telomerase catalyt
27	25	96.2	1480	2	T21911	hypothetical prote
28	25	96.2	1483	2	T21912	hypothetical prote
29	25	96.2	1483	2	T21914	hypothetical prote
30	23	88.5	179	2	E70212	hypothetical prote
31	23	88.5	185	2	C70220	conserved hypothet
32	23	88.5	192	2	G71084	hypothetical prote
33	23	88.5	200	1	H71222	hypothetical prote
34	23	88.5	210	2	G71685	outer membrane ant
35	23	88.5	211	2	G64136	probable DNA-(apur
36	23	88.5	219	2	T25904	hypothetical prote
37	23	88.5	226	2	G45170	cytochrome-c oxida
38	23	88.5	226	2	A42622	H+-transporting AT
39	23	88.5	229	2	H45170	cytochrome-c oxida
40	23	88.5	244	2	S48482	hypothetical prote
41	23	88.5	246	1	PYNVPF	polyhedrin - Panol
42	23	88.5	246	1	A42756	polyhedrin - Mames
43	23	88.5	246	2	A38981	polyhedrin precurs
44	23	88.5	254	2	T41477	hypothetical prote
45	23	88.5	255	2	D69448	conserved hypothet

ALIGNMENTS

RESULT 1
H70070
hypothetical protein ywzC - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: H70070
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: H70070
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-74 <KUN>
A;Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15788.1; PID:e11862
A;Experimental source: strain 168
C;Genetics:
A;Gene: ywzC

Query Match 100.0%; Score 26; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 3 NEFRV 7

RESULT 2
T28997
hypothetical protein ZC513.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28997
R;Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid ZC513.
A;Reference number: Z20551
A;Accession: T28997
A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA
A;Residues: 1-156 <WUX>
A;Cross-references: EMBL:U53155; PIDN:AAC48269.1; GSPDB:GN00023; CESP:ZC513.7
A;Experimental source: strain Bristol N2; clone ZC513
C;Genetics:
A;Gene: CESP:ZC513.7
A;Map position: 5
A;Introns: 95/1

Query Match 100.0%; Score 26; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 129 NEFRV 133

RESULT 3
S39770
seminal vesicle autoantigen precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S39770; S39771; JC5387
R;Yu, L.C.; Chen, J.L.; Tsai, W.B.; Chen, Y.H.
Biochem. J. 296, 571-576, 1993
A;Title: Primary structure and characterization of an androgen-stimulated autoantigen pu
A;Reference number: S39770; MUID:94107218
A;Accession: S39770
A;Molecule type: mRNA
A;Residues: 1-158 <YUL>
A;Cross-references: EMBL:M94179; NID:g191945; PIDN:AAA16593.1; PID:g191946
A;Accession: S39771
A;Molecule type: protein
A;Residues: 28-38,'Y',40-46;73-74,'X',76-86;133-144 <YUW>
R;Yu, L.; Hsiao, Y.; Yang, Y.; Lin, M.; Chen, Y.
Biochem. Biophys. Res. Commun. 231, 106-110, 1997
A;Title: The genomic structure of a mouse seminal vesicle autoantigen.
A;Reference number: JC5387; MUID:97223408
A;Accession: JC5387
A;Molecule type: DNA
A;Residues: 1-158 <YUA>
A;Cross-references: GB:L44117; NID:g914834; PIDN:AAA73510.1; PID:g914835
C;Comment: This protein is a zinc-binding glycoprotein. Its expression is stimulated by
C;Genetics:
A;Gene: sva
A;Introns: 32/2; 77/3; 115/1
C;Keywords: glycoprotein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-158/Product: seminal vesicle autoantigen protein #status experimental <MAT>
F;39,149/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 26; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 60 NEFRV 64

RESULT 4
S28670
avrPph3 protein - Pseudomonas syringae pv. phaseolicola
C;Species: Pseudomonas syringae pv. phaseolicola
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Nov-1999
C;Accession: S28670
R;Jenner, C.; Hitchin, E.; Mansfield, J.; Walters, K.; Betteridge, P.; Teverson, D.; Tay
Mol. Plant Microb. Interact. 4, 553-562, 1991
A;Title: Gene-for-gene interactions between Pseudomonas syringae pv. phaseolicola and Ph
A;Reference number: S28670
A;Accession: S28670
A;Molecule type: DNA
A;Residues: 1-267 <JEN>
A;Cross-references: EMBL:M86401; NID:g151056; PIDN:AAA25727.1; PID:g151057
C;Genetics:
A;Gene: avrPph3
C;Superfamily: Pseudomonas syringae pv. phaseolicola avrPph3 protein

Query Match 100.0%; Score 26; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 262 NEFRV 266

RESULT 5
C70930
hypothetical protein Rv1794 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: C70930
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: C70930
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-300 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17715.1; PID:e125
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1794

Query Match 100.0%; Score 26; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NEFRV 5
|||||
Db 114 NEFRV 118

RESULT 6
T27185
hypothetical protein Y54G9A.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27185
R;Smye, R.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20324
A;Accession: T27185
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-343 <WIL>
A;Cross-references: EMBL:AL032648; PIDN:CAA21698.1; GSPDB:GN00020; CESP:Y54G9A.6
A;Experimental source: clone Y54G9A
C;Genetics:
A;Gene: CESP:Y54G9A.6
A;Map position: 2
A;Introns: 78/3; 291/2

Query Match 100.0%; Score 26; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 43;